

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

WELTORGANISATION FÜR GEISTIGES EIGENTUM
 Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<p>(51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027</p> <p>(21) Internationales Aktenzeichen: PCT/EP98/08216</p> <p>(22) Internationales Anmelde datum: 22. Dezember 1998 (22.12.98)</p> <p>(30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE</p> <p>(71) Anmelder (<i>für alle Bestimmungsstaaten ausser US</i>): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Rogendorfstrasse 59, D-51061 Köln (DE).</p> <p>(74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).</p>	<p>(11) Internationale Veröffentlichungsnummer: WO 99/33998</p> <p>(43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)</p> <p>(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Veröffentlicht <i>Ohne internationalem Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>
<p>(54) Titel: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF</p> <p>(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG</p>	
<p>The diagram illustrates the RT MOTIFS of the human catalytic telomerase sub-unit gene. It shows the gene structure with various mutations. The mutations are categorized into two main types: DELETIONS and INSERTIONS. The mutations are labeled as DELI, DELI-2, DELI-3, INS1, INS2, INS3, and INS4. The diagram also shows the FULL LENGTH of the gene.</p>	
<p>(57) Abstract</p> <p>The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.</p>	

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	MW	Malawi	TR	Türkei
BG	Bulgarien	HU	Ungarn	ML	Mali	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MN	Mongolei	UA	Ukraine
BR	Brasilien	IL	Israel	MR	Mauritanien	UG	Uganda
BY	Belarus	IS	Island	MW	Mosambik	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	MX	Mexiko	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NE	Niger	VN	Vietnam
CG	Kongo	KE	Kenia	NL	Niederlande	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NO	Norwegen	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	NZ	Neuseeland		
CM	Kamerun	KR	Republik Korea	PL	Polen		
CN	China	KZ	Kasachstan	PT	Portugal		
CU	Kuba	LC	St. Lucia	RO	Rumänien		
CZ	Tschechische Republik	LI	Liechtenstein	RU	Russische Föderation		
DE	Deutschland	LK	Sri Lanka	SD	Sudan		
DK	Dänemark	LR	Liberia	SE	Schweden		
EE	Estland			SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten 10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Des Weiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomere enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der 30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

30 Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen 5 auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens 15 (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomerase-aktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et* 20 *al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

25 Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt 30 werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. 10 Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere. 15

20 Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweligen Strahlen schädigen nicht nur die Tumorzellen, sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle 25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden 30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapierung von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-transkriptionale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20
25
30

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements) be-

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

5

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer 10 Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

30

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon beinhaltet.

5

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

10

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15

Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).

20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25

Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

30

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, 5 M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispiels- 10 weise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bak-
terielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase
aus *Secale careale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle
15 Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane
Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phospha-
tase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion
Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthal-
ten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-
Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthal-
ten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen bei-
spielsweise:

30 Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -
Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes
Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein
(GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

5 Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

10 Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

15 Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfundungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

20 Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccinia Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

25 Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfundungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfundungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilesequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20 Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

15 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

20 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;

C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Spraque Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklos P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des 10 Vektors EMBL3 Sp6/T7.

15 Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklos, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

20 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur 25 14: nicht verdaut.

30 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 10 5' GTTAAGTTGTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den 15 Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die 20 dargestellte Sequenz entspricht SEQ ID NO 2.

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25 Exons sind als durchnummierierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymeschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, Xhol. Die relative Anordnung der 30 fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

10

15

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgesplicte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

25

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

30

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

15 Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das pormotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite sind die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

20

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

25

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

30

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
10 Intron 2.

Tab. 1

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der 5 Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

10

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen 15 bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene 20 genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, 25 Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschritte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

30

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 μ m; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 μ g/ml 10 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte 15 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt 20 Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ -DNA Präparation dieses Phagens Ausubel *et al.* 25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen 30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) 5 gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach 10 λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umkloniert (Beispiel 4).

Beispiel 3

15 Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

20 Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

25 Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-
5 Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

10 Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinationsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.
15

20 Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.
25

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenkrons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenkron P17 und 5' von Phagenkron P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl

10

wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

15

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-

20

Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

25

ACTATAAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

30

kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTGAAGGCCTT-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten 5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking 10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen 15 genetischen hTC-Region identifiziert (s. Fig. 7).

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelnsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software 20 (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklobereiche sind in Fig. 7 schematisch 25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

5	ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70	
	ATGAGACCC	GTCTCAAAA	AAAAAAA	AATTGAAATA	ATATAAACGA	TCTTCTCTGG	CCACAGTGG	140	
	ACAAAACCC	AAATCAACAA	CAAGAGGA	TTTGAAA	ACTACAAACAC	ATGAA	ATTA	210	
	TTCTGAATG	CCAGTGACTC	ATGAA	GAAGAA	TTAAAAGGA	TTTATTTAAG	CAATGATAA	280	
	CGGAAACATA	ACCTCTAA	ACCCACGG	TACAG	AAAAA	GAAGAAGTT	TATAGCTATA	350	
	AGCAGCTACA	TCAAAAAGT	AGAAAAGCCA	GGCGCA	CTG	CTGCTG	TAATCCCAGC	420	
	GCCAGGCGG	GCAGATCGCC	TGAGGT	CAGG	AGTCGAGAC	CAGCCTGAC	AACACAGAGA	490	
10	CTACTAAAAA	TACAAAATT	GCTGGCATG	GTG	GGCATG	CCTGTAATCC	CAGCTACTCG	560	
	GCAGGATAAC	CGCTTGAACC	CAGGAGGTG	AGGTTGCGG	GAGC	GGGGAT	TGCGCCATTG	630	
	TGGGTAACAA	GAGTGACTC	CTG	TCAAG	AAAAGA	AAGTGAAGAA	ACTTAAA	700	
	GATGCACTT	AAAAGA	AAAGAAGA	GCAAA	CTTAA	AAATTTAG	ACAAACCTAAT	770	
	AAGATCAGAG	CAGAA	ATAAA	TGAA	AGATAAC	ACAAAAGATC	ACAAAATTA	840	
	TTTGAAA	ATAACAA	TTGACAA	TTG	CCCGAGA	CTAAG	AAAAGGAAG	ACCTTAAATAA	910
	ATAAACTCG	AGAT	AAAAGA	AGACATCA	CAACT	CACAGAA	CAAAGGATCA	980	
	CTATGACCAA	CTG	ACACTA	ATAA	TTGAA	AAACCTGAA	AAATAGATA	1050	
	CTACCAAGAT	TGA	ACATGA	AGAA	AACTCAA	GACCAAA	AATAA	1120	
	AAATAAAAGT	CTCT	AGCAG	AGAGA	GGAC	CCCAAT	GGCTTCC	1190	
15	AAAGAAGAAT	GAAT	TCAT	CCTA	CTATT	CTG	GATT	TTTA	1260
	TCTACATGGC	CA	GAT	TAC	TTG	ACACAT	AAAACAA	1330	
	CA	AAAGAA	AAA	ACTACA	CTG	TGAT	AAAT	CTCAA	1400
	GC	AAAC	AA	TTAA	ACAA	CCTC	GAAAG	ATCATT	1470
	AAGGATGGT	CAAC	ATATG	AA	AT	TGATCA	GGGATT	TTCAGG	1540
20	TATGATTATT	TCAC	TTAG	CAAG	AAAG	TTTG	ATTC	TGAC	1610
	AAAACCAGG	TAT	AAAGA	ACAT	ACAG	CAGG	GGCT	ACAC	1680
	AGGCCAAGGT	GGG	GAT	TGG	GGG	GGG	TGCG	ATCCC	1750
	CTACAAA	AA	TTT	TTA	AA	TTA	CATG	CTG	1820
	GCTGAGGTGG	GAGA	ATC	TAAG	CTAG	AGG	TGCA	GTC	1890
25	CCAGCCTAG	CAAC	AGA	AGA	AGA	GGAG	TGAG	GTG	1960
	AAGGGAGGAG	GGG	AGG	GGG	GGG	GGG	GGG	GGG	2030
	AAGAAACATA	TTT	CAACAT	ATAA	AAAGCC	TATAT	GACAG	ACCGAG	2100
	AGCCTTCC	CTAAG	ATG	GAAC	AGG	GGG	GGG	GGG	2170
	AGTCC	AG	AGA	AA	AA	AA	AA	AA	2240
30	TTATCCGTG	TG	CAG	ATG	ATG	TTA	GACT	AAAG	2310
	GCTGAATTT	GGT	ACAG	CAG	GAT	AA	AA	AA	2380
	AAACAACTG	AAA	AAAG	AAAG	AAAG	AAAG	AAAG	AAAG	2450
	GTGAAAGATC	TCT	ACAT	AA	AA	AA	GCT	AGG	2520
	AGATATTCCA	TG	TTG	AAAGA	AA	AA	AA	AA	2590
35	AAATTCAATG	CAAT	CCCT	AA	AA	AA	AA	AA	2660
	TTG	TACAG	AA	CC	AA	AG	AA	AA	2730
	CA	ATT	AC	CA	AA	AA	AA	AA	2800
	AGATGAGACA	TG	GG	GG	GA	AA	AA	AA	2870
	TTTTGACAA	AGG	TG	CC	AA	AA	AA	AA	2940
40	CTGGATATCC	ATATG	AA	AA	AA	AA	AA	AA	3010
	GGATGAAAGC	CTT	AA	AA	AA	AA	AA	AA	3080
	GGACATTGGA	GT	GG	AA	AA	AA	AA	AA	3150
	AAATGGGATC	ATATCA	AA	AA	AA	AA	AA	AA	3220
45	CCACAAATG	GG	GA	AA	AA	AA	AA	AA	3290
	GCTCAA	ACT	CT	CT	AA	AA	AA	AA	3360
	CATT	CTC	AA	AA	AA	AA	AA	AA	3430
	GAGAAATGCA	AA	AA	AA	AA	AA	AA	AA	3500
	AGGCAAA	AA	AA	AA	AA	AA	AA	AA	3570
50	TTGCTAC	TAT	GG	AA	AA	AA	AA	AA	3640
	CCATTGCTAG	TTA	TAT	AA	AA	AA	AA	AA	3710
	TGCA	GG	AA	AA	AA	AA	AA	AA	3780
	AAATATGTTG	GC	AC	AT	AA	AA	AA	AA	3850
	CAGCATGGG	GG	CT	AT	GG	AA	AA	AA	3920
55	CTCCCTACT	TGT	GG	GG	AA	AA	AA	AA	3990
	GGTGGGAGAC	GG	GG	GG	AA	AA	AA	AA	4060
	TGGGGGAG	AGG	GT	CA	AA	AA	AA	AA	4130
	TGGGGGAG	TT	AA	AA	AA	AA	AA	AA	4200
	TTATTACACA	TT	TG	TG	AA	AA	AA	AA	4270
	AAATTA	AA	AA	AA	AA	AA	AA	AA	4340
60	GTGATCACC	TG	GG	AA	AA	AA	AA	AA	4410
	TGAGGTGAG	AG	TT	AA	AA	AA	AA	AA	4480
	TTGCTTAC	AA	AA	AA	AA	AA	AA	AA	4550
	GAGCAAGACT	CC	AT	CT	AA	AA	AA	AA	4620
	AAATATACT	CT	ACT	AT	AA	AA	AA	AA	4690
65	AAATAAGAAC	AA	ATG	TG	GG	AA	AA	AA	4760
	GTGAGGAGG	AA	AC	AT	GG	CT	AT	CT	4830
	TACAA	AA	AA	AA	AA	AA	AA	AA	4900
	TAATCTTAT	TA	TT	AA	AA	AA	AA	AA	4970
70	CACCGTCTC	TC	ATT	TC	AA	AA	AA	AA	5040
	GTAA	AA	AA	AA	AA	AA	AA	AA	5110
75	TG	GG	AA	AG	AC	CT	TC	GG	

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAGCTC	GGATGGGAAG	GGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTAG	CAGGGCATGA	AGTGCCCTTA	TTAACGCTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
5	AGCCGGCCAG	CGGAAATGCA	AGGAGTCAGA	AGCCTCTGC	TCAAACCCAG	GCCAGCAGT	ATGGGCCCA	5320
	CCCAGGGCTG	TGCCAGAGGG	AGAGGAGTC	AGCCTCTCG	AAGTATGGCT	TAATCTTTT	TTTCACCTGA	5390
	AGCAGTGACC	AAGGTGATT	CTGAGGGAAAG	CTTGAGTTAG	GTGCTTCTT	AAAACAGAA	AGTCATGAA	5460
	GCACCCCTCT	CAAGGGAAA	CCAGACGCC	GCTCTGGGT	CATTACCTC	TTCTCTCT	CCCTCTCTG	5530
	CCCTCGCGGT	TTCTGATGCGG	GACAGAGTC	CCCCCTGGGA	GCTCTGGCGA	GCCCGTGCCTG	AGGACCCCTCT	5600
	TGCAAAGGC	TCCACAGACC	CCCGCCCTGG	AGAGAGGAGT	CTGAGCTGG	CTTAATAACA	AACTGGCATG	5670
10	TGCGCTGGGG	CGGACAGCGA	CGGGGGATT	CAAAGACTTA	ATTCATGAG	TAATTCACAC	CTTCCACAT	5740
	CCGAATGGAT	TTGGGATTA	TCTTAATATT	TTCTTAATT	TCTCAAAATA	ACATTCAGGA	CTGCAGAAAT	5810
	CCAAAGGGCT	AAAACAGGAA	CTGAGCTATG	TTTGCAAGG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880
	ATTTCGCGC	CTAAGTACTT	TTTATTGGTT	TTICATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGGG	5950
	AGAGGCCCTG	GCGGAGGCG	CGAGGAGCTG	CCCCGGGGCA	GCGGAGAGAA	GTCACCCGGG	TGGGAGGCTG	6020
15	ACAGCAGGAC	CACTGAGCTG	CCTCCCTG	AGCTGCCAC	TTGGGCAACG	CGAAGGCCGC	CACGCCGCGT	6090
	GTGACTCAGG	ACCCCCATACC	GGCTTCTGG	CCCCACCCAC	ACTAACCCAG	GAAGTCACCG	AGCTCTGAAC	6160
	CCGGTGAAGAC	GAACATGACC	CTTGGCTGCC	TGCTTCTCC	GGTGGGCTCAA	GGGATAATGAA	GTGGTGTGCA	6230
	GGAAATGGCC	ATGTAATTA	CAAGACTCTG	CTGATGGGG	CCCTTCTTC	CATCATTATT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGGCATGA	CAAAACTCAG	TACAAACACC	6370
20	ACTCTTTTAC	TAGGCCACAA	GAGCAGCGG	CACACCCCTG	ATATATTAGA	AGTCAGGAGA	AGATGAGGCT	6440
	GCTTCAGGC	ACCCAGCTGC	GGTGACAACA	CGCGCTGAAC	AGCTGTTCC	TCTAGACTAG	TAGACCTGG	6510
	CAGGCACTCC	CCCAGATCT	AGGGCCTGGT	TGCTGCTCC	CGAGGGGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGTGCGC	ACACTGAGGC	CAGGGCTCTG	TCCACACCC	CCGGCTCC	GCCTCAGCTT	CTCCAGCAGC	6650
25	TTCTTAACAC	CTGGGTGGGC	CGCTGTTCTG	TCACCTGTC	CACTGTGTC	TGCTCTAGCG	6720	
	ACGTAGCTCG	CACGGTTCT	CCTCACATGG	GGTGTCTGTC	TCTTCTCCCA	ACACTCACAT	CGGTTGAAGG	6790
	GAGGAGATTG	TGCGCCTCCC	AGACTGGCT	CTCTGAGCTG	GAACCTGCGT	CCTGAGGTT	GATGAGGCTT	6860
	CTTGGGCTCC	GGCTGCACCC	TGACCTCTAT	TTCCAGCGC	TCCCCCTCTC	CTGTCATCTG	CGGGGCCCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCTT	TCCACGTC	GCTGCGTGTG	TCTCTGCCG	CTAGGGCTC	7000
30	GGGGTTTTA	TAGGCATAGG	ACGGGGCGC	GGGGGGCTG	GGGGGGCTCA	GGGAAATGCAA	CATTGGGTG	7070
	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGGCTGG	GATGGAGCCC	CCGCCAGGGA	7140
	CCCGCCCTTC	TCTGCCACG	ACTTTCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TCCCACAAAC	7210
	ACTAAGGATC	CTCTCCCAA	AGAGCCACG	ATGGCACCC	CTGGACATT	GCCCCACACC	CCTGGGAAATT	7280
	CACGTGACTA	CGCACATCAT	TGACACACTC	CGCTCACGCA	CGGACCCCCC	CTGTTTATT	TTAATAGCTA	7350
	CAAAGCAGGG	AAATCCCTGC	TAAGATGTC	TTAACAAAC	TGTTAAACA	AAACGGGCTCA	TCGCACGGT	7420
35	GGACAGTTCC	TCACAGTGA	GAGGAACATC	CGGTTTATAA	AGGCTGAGC	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAC	TGCAACCTTC	ATGGGATACC	TACGCACAT	GCTCAAAAG	AAAGAATTTC	ACCCCATGGC	7560
	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGGCGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTACTA	7630
	AAAGCCATTG	CCTGGTTCTG	ATGGTATTGG	CTCAGTTAG	GGAGACTAAC	CATAAGGGAG	TGGGATGGG	7700
	GGAAACCGGA	GGCTGTGCCA	TCTTGGCTAT	GGTGTGAGT	CCTGGGAGG	ATAATGCTCT	AGAGATGCC	7770
40	ACGTCTGTAT	TCCCCAAC	CTGTGAGAC	AACCGCCCG	GGCCCAGGG	CTTTCAGGT	GTGATCTCG	7840
	TGAGGCCCT	GAGGTCTGG	ATCCTCTGCA	GGCCGAAAGA	GGGCTGAAAG	TAATTCAGG	GGTTCTGGGA	7910
	AGAGGGCCGC	AGAGGGTCA	GAGGGGGCGA	GCCCTGAGG	GATGGAGGCA	GTCACTCTG	GGCTGAAAAG	7980
	GGAGGGAGGG	CCTCGAGCC	AGGCTGCA	GGCCCTCCAG	AAAGCTGAAA	AGCGGGGAA	GGGACCCCTC	8050
	ACGGGAGCTC	CAGCAGGAGA	GGACGGCTGG	CCCCTAGGG	ACAGGGGCC	ATCGTGGACC	TCCGGCCTCC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCTT	CTAGCATGAA	GTGTGTTGGG	ATTTCAGAA	GCAACAGGAA	8190
	ACCCATGCAC	TGTGAATCTA	GGATTATTTC	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260
	GTGCCATGGC	GCACTGGCAG	GGCCTGAGG	AGTGTGTTTA	TTAGTCTT	TTATTTTATT	TACTTACTTT	8330
	CTGAGGACAGA	GTTATGCTCT	TTTGGCCCT	GTGTTGAGTC	AGGGCGATGA	TCTTGCTCA	CTGCAACCTC	8400
	CGTCTCTGG	GTTCAAGCAA	TTCTCGTGC	TCAGCCTCCC	AAAGTAGTGG	GATTTCAAGG	GTGCAACACC	8470
50	ACACCCGGCT	AATTGTTGAT	TTTGTAGTA	TGATGGCTAT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATGCCG	ACACCTCGCC	TCCCAAGAGT	CTGGGATATC	AGGCATGAGC	CACTGCACCT	8610
	GGCCTATTAA	ACCATTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAGGGAGTT	CATGGAGTTC	8680
	AAATTCCCTC	CTACTCAGGA	GGTACCTCTC	TTTGTGATTT	TCTGTAAT	TTCTGAGACT	GGGGATACAC	8750
55	CGTCTCTGTA	CATATTCA	TTTCTGTGA	CCACCTGTTA	TCCCCATGGGA	CCCACTGCG	GGGCAGCTGG	8820
	GAGGCTGCA	GCTTCAGGTC	CCAGTGGGGT	TGCCCATGTC	CAGTGAACAC	CTGATGTTGA	ATCAGGGCG	8890
	AACTGTGCGAC	ACTGTCCTGA	ATCTCAATGT	CTCAGTGTG	GCTGAAACAT	GTAGAAATTAA	AACTCCATCC	8960
	CTCTTACTCT	ACTGGGATTG	AGCCCCCTTC	CTACCTCCCC	CGAGGGCGAC	AGGAGTTCT	CTCCTCCCTG	9030
	TGGAGGAAGG	ATATGAACTT	TGTTATTCTT	CACTGCTGT	ACTGAATCA	CTGTTTATT	TGTTGTTTG	9100
	TTTGTGTTGT	TTTGTGAGGG	GGTTTCACTG	TTGTTGCTCA	GGTGTGGAGG	AGTGAATGG	GGCGATCTTG	9170
60	GCTTACTCTGCA	GGCTCTGCT	CCACCATGCC	CACTAATT	TTTGTGTTT	TAAGTAGAGAC	GGGGTTGGGT	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATT	TTTGTGTTT	TAAGTAGAGAC	GGGGTTGGGT	GGGTTTACCC	9310
	ATCTTGGCCA	GGCTGCTGTC	AACTCTCTG	TTCCAGATGA	TCCACCTGCG	TCTGCTCTCT	AAAGTGCTGG	9380
	GATTACAGGT	GTGAGCCACC	ATGCCCACTG	CAAGAATTAC	TCTGTGTTAGA	AACTCTGGG	TCTGAGGTAG	9450
	GAAGCTCACC	CCACTCAAGT	TTTGTGTTGT	TTAACGCAAA	TTGATAGAA	TTTTTATTGT	TGTTAGAA	9520
65	CTCTTGTGAT	TTTACACTG	ATGATGAGC	ACATCATCAG	CTTTTCAAG	AAACACTAAC	TGCAACCCATA	9590
	ATACTGGGGT	GGTCTCTGGG	TATCAGCAA	CTTCATGGA	TTGCGGGAGG	CGTTTCTCG	CCATGCACAT	9660
	GGTGTGTTATT	ACTCCAGCAT	AATCTCTGC	TTCCATTCT	TCTCTTCTCT	CTTTTAAAT	TGTGTTTCT	9730
	ATGTTGGCCTT	CTCTGCAAGAG	AACTGAGTC	AGCTAACACT	TAACTTTTGT	TGGAACAAAT	TTCCAAACC	9800
	GCCCCCTTGC	CCTACTGGCA	GAGACAATT	ACAAACAGC	CCCTTTAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTAG	AGAGCGACCC	TGTAATCCCA	AGTATTAA	AGACGAGGCT	AACTCCAGC	GAGCGTGA	9940
	GCCCCAGGGAG	GGTGGAGGAG	CTGTTCAAAT	GTAGGTCTCA	TAATATAAGC	AATTTCCTCC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAAGTTTACAT	TTAACGTTGC	TTTGTGTTAGC	ATTTCAAGT	TTGCGGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCC	CCTTAGATC	AAACTTGAGC	AAACCCGGAGT	10150
	CTGGGATTCT	GGGAAGTCTT	CACTGTCCT	GGGGTTGTG	CGGGGGCCCA	GGTCTGGAGG	GGACCAAGTGG	10220
	CCGTGTGGCT	TCTACTGCTG	GGCTGAGGT	CGGGCTCTCT	AGCTCTGAG	TCCGAGGCTT	GGAGCCAGGT	10290
75	GCCTGGACCC	CGAGGCTGCC	CTCCACCTCTG	TGCGGGCGG	ATGTGACCG	ATGTTGGCT	CATCTGCCAG	10360
	ACAGAGTGCC	GGGGCCCAAGG	GTCAAGGGCC	TTGTGGCTGG	TGTGAGGGCC	CCGGTGC	GGCCAGCAGGA	10430
	GGCCTGGCT	CCATTCTCCA	CCCTTCTCTG	ACGGGACCCG	CCCGGGGGGT	GATTAACAGA	TTTGGGTTGG	10500

5 TTTGTCTATG GTGGGGACCC CTCGCGGCC GAGAACCTGC AAAGAGAAAT GACGGGCCCTG TGTCAAGGAG 10570
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCGT GCCCCTCCAG GGAGCAATGC 10640
 GTCTCTGGGT TCGTCCCCAG CGCGCTCTAC GGGCCCTCGG CTCCTCCCTC ACGTCCGGCA TTGCTGGTGC 10710
 CCGGAGCCCG AGCCCCCGC TCCGGACCTG GAGGCAGCGG TGGGTCCTCG GATCAGGCCA GGGGCCAAG 10780
 GGTGCGGCCA CGCACCTGTT CCCAGGGCTT CCACATCATG GCCCCCTCCCT CCGGTTACCC CACAGCTAG 10850
 GCGGATTCTGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCCT GGGAGCGCGA GCGGCGCG 10920
 GGCCTGGGAGG CGCGGCCCAAC ACCCCCCGGT CGCGGGCGGAG CAGCTGGCGA GTCGGGGCCA GCGGCGCG 10990
 CCAGTGGATT CGCGGCCACCA GACGCCAGG CGCGCTCTCC ACACGTGGCC GAGGGACTGG GAGACCGGGC 11060
 10 ACCCGTCTG CCCCCCTCAC TCCAGCTCC GCTCTCTCCG CGCGGACCCC GCCCCGTCCC GACCCCTCCC 11130
 GGGTCCCCGG CCCAGGCCCT TCCGGCCCT CCCAGGCCCT CCCCTCTCC TCCGGGGCCC CGCCCTCTCC 11200
 TCGCGGCCG AGTTTCAGGC AGCGCTCGCT CGCTGGCGC ACCTGGGAAG CGCTGGCCC GGCACCCCC 11270
 GCGATGCCGC CGCGCTCCCG CTGGCGAGCC GTGCGCTCCC TGCTGCAGG CCACTACCGC GAGGTGCTGC 11340
 CGCTGGCCAC GTTCTGGGAGG CGCCCTGGGG CCCAGGGCTG CGCGCTGGCAG CAGCGGGGG ACCCGGGG 11410
 TTTCCGGCG CGTGGGGCCC AGTGCCTGGT TGCGCTGGCC TGGGACGCAC GGCACCCCCC CGCCGCCCC 11480
 15 TCCCTCGGC AGGTGGGCT CCCCGGGTC GCGCTCCGC TGGGTTGAG GGCACGGGG GGAACCAAGC 11550
 GACATGGGA GAGCAGGCA GGCAGACTAG GGCCTCCCG CGCAGGCTTCT CTCGCTGAA GGAGCTGGT 11620
 GCCCCGACTGC TCGAGGGCT AGTGCAGCGC CGCGCGAAGA AGCTGGCTGC CTTCGCTTC GGCCTGCTGG 11690
 ACAGGGGCCG CGGGGGCCCG CCCGAGGCCCT TCAACCACAG CGTGCAGCAG TACCTGCCA ACACGGTGAC 11760
 CGACGACTG CGGGGGAGGG CGGGCTGGCTG CGCGCTGGC AGCAGGACCT GCTGGTTCAC 11830
 20 CTGCTGGCAC GCTGGCGCT TTTTGCTG TGCGCTCCCA GCTGCCTAAGA CAAGGTGTC GGGCCGCC 11900
 TGTAACAGCT CGGCCTGCGC ACTCAGGCC GGCACCCGCC ACACGCTAGT GGACCCGAA GGCCTGCTGG 11970
 ATGCGAACGG GCCTGGAAAC ATAGCTGAG CGGACGGCCG GTGGGGCTGG CGCTCCGAGC CGGGCTGGC 12040
 AGGAGGGCGG GGGCGAGCTG CAGCGGAGT CGCGCTGGTCC CAAAGAGGCC CAGGGCTGGC GCTGGCCC 12110
 AGCCGGAGCG GACGCCCGT GGGCAGGGT CCTGGGCCA CCGGGCAGG ACGCCTGGAC CGAGTGACCG 12180
 25 TGGTTCTGT GTGGTGTAC CGCCAGGAGA GCAAGCTCTT TGAGGGTGC GCTCTCTGCG 12250
 ACAGCGCACT CCCACCATC CTGGGGCCG CAGCACACAG CAGGGCCCCC ATCCACATCG CGGCCACAC 12320
 GTCCCTGGGA CACGCTTGT CCCCCGGTGT ACGCCGAGAC CAAGCACTTC CTCTACTCTT CAGGCAGCAA 12390
 GGAGCAGCTG CGGGCTCCCT TCTCTACTCG CTCTCTGAGG CCGACCTCTA CTGGGGCTCG GAGGCTCGT 12460
 GAGACCATCT TTCTGGGTG CAGGGCTCGG ATGCCAGGA CGCCCGAGGAG CAGTGCCTCC ACGGGGTGT 12530
 30 GCTACTGGCA ATAGGGGCC CTGTTCTGG AGCTGCTGG GAACCCAGG CAGTGCCTCC ACGGGGTGT 12600
 CCTCAAGACG CACTGCCCGC TGCAGGCTG CGGCACCCCA CGACGGCGGT TCTGTCCTGG GGAGAAGCCC 12670
 CAGGGCTCG TGCGGGCCCG CGAGGGAGAG GACACAGAC CCGCGCTCGG GGTGAGCTG CTCCGCGAC 12740
 ACAGCAGCCC CTGGCAGGTG TACGGCTCG TGCGGGCTG CCTGCGCCG TGTTGCCCC CGGGCTCTG 12810
 GGGCTCAGG CACAAACAC GCGCTCCCT CAGGAACACC AAGAAGTICA TCTCTCTGGG GGAGCATGCC 12880
 35 AAGCTCTCGC TGCGAGGCT AGCGCTGAG ATAGCGCTGC GGGACTGCGC TTGGCTGCCA AGGAGGCC 12950
 GTGAGGAGGT GGTGGCCGTC GAGGGCCAG GCCCCAGAGC TGAATGCAGT AGGGGCTCA AAAAGGGGG 13020
 AGGCAGGCC CGTGGCTCTC TGCTCCATC GTCACTGAGG CACAGTGGT TTTGCTCTCA GGACGCTGAG 13090
 TGGACACGGT GATCTCTGC TCTCTGTCCA GTTGTACATA ACTTACAGGG TTACCTTCA 13160
 40 CGTTTGATG GACACGCCGT TTCCAGCGC CGAGGCCAGA GCAGTGAACA GAGGAGGCTG GGCACGGCAG 13230
 TGAGGGGGGTT TGCGGGCAA TGAGGAGAAG CACAGACCCG CTGGGGAGGG TGCTGGCAGG 13300
 TTACCTATAA TTCTCTCCG AATTCAAGG TGCGGAATGA GAGGGGGGA CGAGAACCCC CTCTCTGG 13370
 GGGTGGGAGG TAAGGGTTTG CGAGGTGAC GTGGTCAAGC AATATGCAGG TTGTGTTTA AGATTAATT 13440
 GTGTGTTGAG CGCCAGGCTC CGGGCTCAC CGCCACTT GGGAGCTGAG GGAGGGTGA 13510
 45 TCACCTGAGG TCAAGGAGTTT GAGACAGCCG TGACCAACAT GGTGAACCC TATCTGACT AAAATACAA 13580
 AAATTAGCTG GGCATGGTG TGTGCTCG TAATCCAGC TACTTGGAG GCTGAGGAG GAGAACACT 13650
 TGAACCCAGG AGGGCGAGGC TGCACTGAG TGAGATTGTC CCATTTGAGT CCAGGCTGGG CGACAAAGAGT 13720
 GAAACACTGTG CTTTAAAAAA AAAAAGTGTG CGTGGATTGGT GCGAGGACAG GTAGGGAGA GGGAGATAAG 13790
 ACTGTTCTCC AGCACAGATC CTGGTCCCAT CTTAGGTAT GAAGAGGGCC ACATGGGAGC AGAGGACAGC 13860
 AGATGGCTGC ACCTGCTGA GAAGGGACAG TGTTTGAGG TGTTGAGGG ATGGGCTGCC TGGGGCCCTG 13930
 50 CGTGTCCCCA CCGTGTGTTT CTGGATTGA GTTGTAGGAA CCTCGCTCGG ATGGGCTTAA CGCTCATGCA 14350
 TGCTCCAGG CCTTACCGTG CGAGCTGAA GAAGTCCCGA TTTCACCCCC TCCCCACAAA CTCCAAGAC 14070
 ATGTAAGACT TCCGCGCATG CAGACAGGG GGGTGCCTG CTGGGGCTC TTGTTTCTTCTT TTGTTTCTT 14140
 TTATGGTGGC AAAAGTCATA TAAACATGAGA TTGGCACTTC TAAACCGTT TCTGTGTC AGTGCAGAT 14210
 TGCTAACTCG GCGGTGTTA CAGCAGGTTG CTTGAATGC TGCGTCTGC GTGACTGGAA GTGCCCTACCC 14280
 ATGCAACGCCG AGCTGCTCA CACTCTGCTG CGCTCAGGG GACCGCCCG AGTCAGATAA CGCTCATGCA 14350
 ACCAGGAGTT GCTTTTTG TGCTCGCTT CTTGGTTGA GAGAGTTGA GTTCTGTAT CAGGACTCTG 14420
 CCTGTCTTGTG CTGTTCTCTG ACTTCAGATG AGGTCAACAT CTGGCCCTGG CTATGAGG GAGTGGAGGG 14490
 TGTTCTCCCGG GTGCTCTCTG CAGCTGAGG GTAGTGGCTG CGTGGCCCTT AGGTGCTCCC GTACAGTGT 14560
 60 GGGTGAGTGA CGCGCGGCCCG CGGGGTGTC CTGTCCTGG CAGCGTGTATT GAGGTGTCGC CCCGGGTGT 14630
 CCTGTCTACG TGAGGGTGA GTGAGGGCCG ATCCCCGGGT GTCCCTGTCA CGTGTAGGGT GAGTGGAGGG 14700
 TGTTCTCCCG GTGCTCTCTG CGGGCTGAGG GTAGTGGCTG CACTGCTCCC GGGTCTCCC GTACAGTGT 14770
 GGGTGAGTGA CGCGCGGCCCG CGGGGTGTC CTCTCAGGGT TAGGGTGTAGT GAGGGCGGG CCCAGGGTGT 14840
 CCTGTCTACG TGAGGGTGA GTGAGGGCACC GTCCCTGGGT GTCCCTCCCA GGTATGGGT GAGTGGAGCA 14910
 CTGTCCTCTG GTGCTCTCTG CAGCTGAGG GTAGTGGCTG CGGCGCCCGG GGGTCTCCC CTACAGTGT 14980
 GGGTGAGTGA CGCGCTGTC CGGGGTGTC CTGTCCTGG TAGGGTGTAGT GAGGGCTGTG CCCGGGTGT 15050
 CCTTGGCGTT TGCTCACTTG AGCTTGTCC TGAATGTGG CTCTTCTAT AGCCACAGCT GCGCCGGTT 15120
 CCCATTGCTC GGGTAGATGG TGCAAGGGCA GTGCTGGTCC CCAAGCTAT TTGTTCTGAT GCTGGCTCT 15190
 TCTGGTAC TCTCTGGTC CATTCTGCTA CGGGGACAGC GGACTGCAAGG CTCTCGCTC CGCGCTGCCA 15260
 70 GGCACTGCA CGACAGCTTC AGGTCCGTT GCTCTGTG TGCTCTGTT GCTCACCACG TGCCCGCCAC 15330
 ATGCATGCTG CCAATACTCC TCTCCCAAGCT TGCTCTCATGC CGAGGGCTGGA CTCTGGGCTG CCTGTGTCG 15400
 CTGCCACCTG TGCTGGAGA CATCCAGAA AGGGTCTCTG TGCCCTGAA GAAAGACAG TCACCCACAG 15470
 CCCCTCACTT TGCTGTGTT CTCCCAAGCT GCCCCCTCTG TTGGCCCTCTT TGGGGGGTT GCAACGCTTG 15540
 TCACCTTATT CGGGCTCATT CGGGCACCT GTCTAGGGTG GGTCTGCCCT CGAGTGGCCC CCTCACATGG 15610
 ATTGACGTCG ACCCACAGGT TGAGTGTCT CTGTCCTGCT CTCGCTCTGA GACCCAGTG GAGGGCCGGT 15680
 75 GTCTCCGCA GCCTCTGTC GACTCCCTC TTGGGTCTTA GTTTGAATT TCACTGATT ACCTCTGACG 15750
 TTCTATCTC TCCATGTTAT GCTTTTCTT CTGCTCTGTT TTCTAGCTTC TTAGTTTGT 15820
 CATGCCCTTC CCTCTAAAGTG CTGCGTCTAC TGACCCCTGT GTTTGTAGT GAAGTAATCT CAACATCAGC 15890

5	CACTTCAAG	TGTCTCTAA	ATACTTCAA	GTGTTAATAC	TTCTTTAAG	TATTCTTATT	CTGTGATTT	15960
	TTTCTTGTG	CACGGCTGTG	TTTCACTGTA	AATCTATTG	ATATCAGTG	CTTTAAAGT	TTCTTCTAGC	16030
	TATTCTGTG	TTTCTTGTG	CACTGACTT	TTTGAACACT	TTTATGTTG	AAGATATGTA	GAGATCTAACG	16100
	ATACGTAGAG	TATTTTAAGT	TATCATTTA	TTATTGATT	CTAACTCAGT	TTGTTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TAAGAATTG	GGAGCCTTGC	TTTGTGATCT	ATGTCATGTC	GGGTTTCCAG	AACTGTCCAT	16240
	TGTAATTATT	ATACCTGTC	AAATAGGGC	ATGATGTTTC	ACTATATCCA	GCTTATTAAAG	GTCTTCTGCA	16310
	AAGCTTCTGT	CTCCCTCTAG	ATGATGAA	TTCCAAGAG	GAGGCCATAG	CTCCCTACCT	GGGGGATGCG	16380
	TCTGTGATT	TCTTCTGTG	TGGTAGCTT	TATGTGAGGC	ATTGTTAGGT	GCATGACGCT	GGTAGAATT	16450
10	TTATCTCTT	GATGAGTGA	TCTTTGGAG	ACTTCTATG	CTCTAGTAAT	CTAGTAATT	TTTTTTAAAG	16520
	TTGCTCTTAG	TACTGCCAC	CTGGGCTCT	TTTGTAGTAGT	TTTCTCTTCG	TGTGTCTGTT	TTCTGCCTT	16590
	AATTATATA	TATATATAT	TTTTTTTTT	TTTGTAGAC	GAGCTTGGT	CTGTGCCCGA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAATT	TACCTTCTGG	CTTGAGCCCT	CTCTCAGCT	CAGCCCTCTG	16730
	AGTACGTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTAAAT	TTTTCTGGG	GAACGGCTC	16800
	TGCTGTGTT	CCCAGGCTGG	TCTCAAAC	TTGGAACCAA	GGGATCCATC	TACCTCGGGCT	TCCCCAAAGTG	16870
	CTGAATTACA	GGCATGAGCC	ACCACGTC	GCCTAACTT	CAACACTT	ATATCTTAT	AGTGTGGGTA	16940
	TGCTCTGTT	ACAGCATGTA	GGTGAATTTC	CAACCTGTC	TCAGACGCTG	TCTTAACTG	GATAACCTGA	17010
15	TTTATTTCA	TTTTTTGTC	ACTAGAGACC	CGCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTGTCATGT	17080
	CCTCGTCTCC	TGTTTCTCA	CCACCTCTG	GGTGTGCGAT	TGCCCTTCTC	GGCCGAGTGC	TGTGTCATCT	17150
	CTCGTGTGCT	CTCGTGTACT	GGGCAATTG	TTTTTTCT	CTTTCCTTGT	TGTTACCCCC	TGATCTTTT	17220
	ATTGTCGTTG	TTTGTCTTTG	TTTATTGAGA	CAGTCTACT	CTGTGACCCCA	GGCTGGAGTG	TAATGGCAC	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCGCTCTC	GGTTCAGGCA	GTTCATTC	CTCAACCTCA	TGAGTACGTC	17360
	GGATTACAGG	GGCCCAACAC	CAAGCCCTGC	TAATTGTTG	ATTTTAAAGT	GAGATAGCTG	TTACCATGAT	17430
	TGGCAGGCT	GGTCTAACAC	TCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
	ACAGGTGCAA	GGCCACCTGG	CGGGCATACC	TTGATCTTTT	AAAAATGAAAGT	CTGAAACATT	GCTACCTTGT	17570
	TCCTGAGCAA	TAAGACCCCT	AGTGTATT	AGCTCTGGC	ACCCCCACAG	CTGTGCTGCT	TTTCTCTTCG	17640
	TGACTTAGT	CTATCTCAGG	CATCTTGACA	CCCCCACAG	CTAACGATTA	TTAATATTGT	TTTCCGTTGTT	17710
	GAGTGTCTT	TAGCTTTGC	CCCCGGCTG	CTTTTCTTC	TTTGTCTCCC	GTCTGCTCTC	TGTCATCGGG	17780
	CCGGCGTCTG	GGGTCCTCTT	CTCTGCTCTT	TGGCTGGTTC	TTCTGTCTG	TTATTGCTG	TAACACCCAG	17850
20	CTTACCTGT	GTGGCCTCC	ATGGCATCTA	GGCAGCTCG	GGGACCTCTG	CTTATGATGC	ACAGATGAG	17920
	ATGTGAGAC	TCAGGAGAC	GGGGCCATTC	TTGGGCTG	AGTGTCTGGA	GCACCCAGCT	GCCAGCTCTG	17990
	CTTACCGAGT	GAGTGACAGC	AACCTCCCT	GGGGCTGGGT	TCAGCGTGG	CTGGCTGGG	ATCATGCGGG	18060
	TCTGGTGGT	CCCGGGTGTG	GAGTTTGAAA	TCGGCAGAAC	CTGGCGTGTG	GGGCCAGCTC	TGACGGTGC	18130
	GCCTGGGGG	GGAGTGTCTG	CTTCCTCTCC	TCTGCTGGG	ACACCGAGCA	AAGGATGAG	CTCCGAGGCG	18200
	TTGTCGCCCC	ACAGGAGCAT	GACGTGAGCC	ATGGTATAAA	TTTTAAAATT	TCTAGGTG	GGCGGGTGGC	18270
	TCACGCCCTG	AATCCCAGCA	CTTGGGGAGG	CCAAGGGGG	TGGATCACGA	GGTCAGGAGG	TCGAGACCAT	18340
	CCTGGCAAC	ATGATGAAAC	CCATCTGTA	CTAAACACAC	AAAAATTACG	TGGCGTGTG	GGGGGGTGC	18410
	TGTAATCCCA	GCGTACTCGGG	AGGCTGAGGC	AGGAGATTG	CTTGAACCTG	GGAGTTGGAA	TTGCACTGTA	18480
	GGCCACATTG	CACCACTGCA	CTTCAGCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAGAA	18550
	AAAAAAAAT	AATTCTAGTA	GGCACCAATTAA	AAAATGAAA	AAAAGAAAAGT	GAATAATTG	TAATAATAGA	18620
30	TTTACTGAA	GCCCCAGCATG	TCCACACCTC	ATCATTTAG	GGTGTATTG	TTGGGAGCAT	CACTACAGG	18690
	ACATTTGACA	TTTTTGAGC	TTTGTCTCG	GGATCCCCTG	TGTTAGGTCCC	GTGCGTGGCC	ATCTCGGCC	18760
	GGACCTGCTG	GGCTTCCCAT	GGCCATGGCT	TTGGTACACC	ATGGTCTGAGG	TCGGGGATGC	GGTGGCTGG	18830
	CCCTCACTG	GCTGGATGTC	CAGTGTCTGG	ATGGTCTGAG	TTGGGATGAT	GGTCAGGCTG	GGGGGGAGG	18890
	AGCTGGATGT	GTGGTGTCTG	GATGGTGAG	TCAGGGGGTG	AGGTCTCCAG	GGCCCTGGTG	AGCTGGAGGT	18970
	ATGGAGTCGG	GATGATGCGAC	GTCCGGGGTG	AGGTGGCCG	GGCCCTGCTG	GAGCTGGATG	TGTTGTTCTG	19040
	GGATGGTGC	GGTCAAGGGT	GGGGCTGCTG	AGGTCAGCTG	AGGTGCGGAG	GGCTCTGGT	GGTGAATGCA	19110
	GGTCCGGGGT	GAGGTGCGCA	GGGGCTGCTG	TGAGCTGGAT	AGGCTGGAGG	TATGGACTCC	GGGATGTC	19180
35	TGAGGTGAC	AGGGCTGCTG	GGTGTCTGGG	TTGGTACACC	TTGGGATGTTG	GGGGGGTGC	GGGGGGGGG	19250
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19320
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19390
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19460
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19530
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19600
40	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19670
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19740
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19810
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19880
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	19950
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20020
45	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20090
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20160
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20230
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20300
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20370
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20440
50	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20510
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20580
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20650
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20720
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20790
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20860
55	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	20930
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21000
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21070
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21140
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21210
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21280
60	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21350
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21420
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21490
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21560
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21630
65	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21700
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21770
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21840
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21910
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	21980
70	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22050
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22120
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22190
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22260
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22330
	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22400
75	GGGCTGGCT	GGGCTGGCT	GGGCTGGCT	GGGGCTGGCT	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	22470

	TGCGGGTGAGC	TGGATGTGTG	GTCTCTGGAT	GCTGCAGGTC	CGGGGTGACT	TGCCCAAGGCC	CTCGGTGAGC	21350
	TGGATATCGC	GTGTCCCCGT	GTCCGAATGG	TGCAAGGTCCA	GGGTGAGGTC	GCCAGGGCCCT	TGGTGGGCTG	21420
5	GATGTGCGGT	GTCCGGATGG	TGCAAGGTCTG	GGGTGAGGTC	ACCAAGGCTC	CGGTGATCTG	GATGTGCGGT	21490
	GTCCGGATGG	TGCAAGGTCTG	GGGTGAGGTC	ACGACCGTCT	CGGTGAGGAG	ATCCTGGCCA	AGTCTCTGCA	21560
	TTAAGGGGT	TGGCTGTGTT	CCGGCCGCAG	TGCTCGAGCT	TTCCTTTATG	TACCGGAGAC	CACCTTTCAA	21630
	CTGGCTGATG	AGTGTGTACG	TGCTCGAGCT	TGCTCGAGCT	AGTGGACAAAG	CATTGGAATC	AGGTACTGTA	21770
	AAGAACAGGC	TCTTTTCTA	CCGGAAGAGT	TGCTGGACCA	AGTGGAAAG	CAGCCCGGCC	TCAGCATGCC	21840
10	TCCCCACGCC	AGGCCCTCTG	TTCTCGAACT	CTCTGGAAAC	TCAGCATGCC	CCTGCTCTCA	21910	
	CTTGGCTCTG	CTTCCCTGGC	TGTCGACGCTC	TGGGGCTGGG	GCCAGGGGCC	CCGTCACAGG	CCTGGTCCAA	21980
	GTGATTCTG	TGCAAGGCTC	TGACTGCTG	GAGCTCACGT	TCTCTTACTT	GTAAAATCG	GAGTTTGTG	22050
	CAAGTGGCT	CTAGGGTTG	TAAGCAGAA	GGGATTAAAA	TTAGATGAAA	ACACTACAC	TAGCTCTCTT	22120
	GCCCTTCCCT	GGGATGTGGC	TCTGATCTC	TCTCTTACTT	TTTGGATGATG	AGTCTCACTC	22190	
15	TGTTGCCAG	GCTGGAGTGC	AGTGGATAAA	TCTGGCTCA	CTGCAACCTC	ACACCTCTGG	GTAAAGCGA	22260
	TTCACCAAGC	TCAGCCTCTC	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCCCTGGC	AATTCTTGT	22330
	CTTTTAGGAG	TCACCATGGT	GGCAGGAGT	GTCTGCAACT	CATGACCTCA	GGTGATCCAC	22400	
	CCACCTTGGC	CTCCCCAAAGT	GCTGGGTTTA	GACCTGTCAG	TCTCTTTAA	GGATAAAATC	CCACCCACTT	22470
20	TTCATGCTGT	TCTGTATGAA	TCTTCATCT	ATTGGATTAA	GGTCATGAGA	GGATAAAATC	TGAGCTTAAC	22540
	GGCGACTCAC	TGCAAGGGAGG	ACCTGTCGAC	GGGAGGACCTG	GGGATAGGAG	AGTTCACCCA	TGAGCTTAAC	22610
	TCTAGGTGGC	TGCAATTGAA	TGGCTGTGAG	ATTGGCTCTG	CAATGTCGG	CTGATGAGAG	TGTGAGATTG	22680
	TGACAGATT	AAGCTGGATT	TGCACTAGTG	AGGGACGGGA	GGCGTGGCT	GGGAGATGCC	AGCCTGGCTG	22750
	AGCCCAAGGC	ATGGTATTAG	CTTCTCTGGT	TGCGGCCCCG	GCTGACTGTG	GAGGGCTTTA	GTACAGAAGAT	22820
	CAGGGCTTCC	CCAGCTCCCC	TGACACTCG	AGTCCCTGGG	GGGCGTTG	ACACCCCATG	CCCCAAATCA	22890
25	GGATGTCTG	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAAACACAGC	CTCTGGCTG	GGGACCCCGA	22960
	CGTGGTGTGG	GGGCCATTTC	CTTGTGCTG	GGGGAGGGTC	AGGGCTTTCC	CTGGGGAAAC	AAGTTAATAC	23030
	ACAATGCAAC	TCTACTGAGC	TCTACAGTA	TTTAAATGGT	TGGCAGCCAA	CAATGTCATT	TGACCAAGTAT	23100
	TTTGGAAAGA	ATTTAATTGG	GGTACCGGA	AGGAGCAGAC	AGACCTGGTG	GTCCCAAGA	TGCTCTTGT	23170
	CACTACTGGG	ACTGTGTTC	TCGCTGGGG	GGCTTGGGAG	CCCCCTCTCC	CTGGACAGGG	TACCGTGCTC	23240
30	TTCTACTCT	GCTGGGCTCTG	CCGGCTTCCGG	TCAGGGCACC	AGCTCCGGAG	CCCCCGGCG	CCCACTGTCC	23310
	ACGGAGTGC	AGGCTGTCTG	CCACAGATGC	CCAGGTCAG	CTGTCGGCC	CTAGCCCC	GTGCCCCCAT	23380
	GGGTGGGTT	GGGGGAAAGA	GCCAAGGGCA	GAGGTGTCAG	GAGACTGGT	GGCTCATGAG	AGCTGATTCT	23450
	GCTCTTGGC	TGAGCTGCC	TGACGACCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTG	CTCTTCTAC	23520
	CTGGGGTCC	TGCCCTGGGG	CAAGCTTGGG	CTACCCCGT	GGCTGTACCA	GAGGGACAGG	CATCTGTGT	23590
35	GGAGGGGCAT	GGGGTCACTG	GGGGCTTGGG	CGGCACTGGG	ACCAGCTGGG	CTGGTCTGGA	TGGTGGGACA	23660
	GTCACCTTGG	GGGGTACCGG	CGGCACTGGG	CGTCCCGGAG	GTTGACTATA	GGACAGGTTG	TCCAGGTGCC	23730
	CTGCAAGTAG	AGGGCTCTC	AGAGGGCTCT	GGCTGGCATG	GGGACAGCTG	GGCCCTTCAG	TGGCTCTCCC	23800
	CGTGTGCTGC	CGTGGGTGCG	CTAGGCCCCCT	ACTGAGTGTG	TGGGGCTTGT	TGAGCTTCCC	GAACGACTT	23870
40	CCTAGTGTG	TCTCTGGCTG	AGCAAGCTC	CTGAGGGGCT	CTCTATGGCA	GAAGAGGGT	GAAGAGGGTG	23940
	CAGCTGCGGG	AGCTGTGCGG	AGCAGAGGTC	GGGAGGATC	GGGAAGCCAG	GGCCGCCCTG	CTGACGTCCA	24010
	GACTCCGCTT	CATCCCCAAG	CGCTGACGGG	TGCGGGCGAT	TGTGAACATG	GACTACGGT	TGGGAGCCAG	24080
	AACGTCGCC	AGAAGAAAAGA	GGGTGGCTGT	GCTTTGGTTT	ACCTCTTCTT	TTAACAGAAA	GTGCGTTTG	24150
	GCCCCACATT	TGTTATCAGC	TGAGATGAG	GGGCCGGAGG	GGGGCCACAG	GGACACAGG	AGGGCCATGG	24220
45	CACGGCGGCC	ACCCATTGTG	GGCCACAGTG	AGTGGCCGA	GGTGGCTG	CTTCCAGAAA	AGCAGCTG	24290
	GGGTGGTAGG	GGAGCTCTGG	GGGAGCTCTG	GGGAGGAGGAC	AGGCTCTGAG	GACCAACAAGA	AGCAGCCGG	24360
	GATGAGCAC	GGCCCGAGGT	CTTGGATCG	TGTCCTGCTG	TGGTGGCGAG	CTCCCTGCG	CTTCGGCTTA	24430
	GGGGGCCCCG	GGACCAAGGG	ACGACTGCCA	GGAGCCCCACC	GGGCTCTGAG	GATCTGGAC	CTTCCCCCAC	24500
	GGGCTCTGCA	CCCCACCCCT	GTGGCTGCC	TGGCTGCGGT	GACCCCTGCA	TCTGAGGAGA	GTGCGGGT	24570
50	AGGTGGACAG	AGGTGTGGC	TGAGGATCCC	GTGTCACAA	CAACATGCCG	CAGGAACCCC	TTTCAAACAG	24640
	GGTGTGAGGA	AGCTGGGAGG	GGTTCTAGGT	CCCCGGCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24710
	TCTCCCCCTG	GTCCCTATGG	TGGGGTGGGG	ACTTGGCCG	ATCCATGGCT	CTGACTGTCT	CCATGCTGT	24780
	CCCCCCCAGG	CGAGGGCTCT	CACCTCGAGG	GTGAGGGAC	TGTCAGGCT	GCTCAACTAC	GAGGGGGCGC	24850
55	GGGGGCCCCG	CTTCCCTGGG	GGCTCTGTG	TGGGCTGTGA	CGATATCCAC	AGGGCTTGGC	GCACCTTCTG	24920
	GCTGGTGTG	GGGGGCCCCG	ACCCGGCCG	TGAGCTGAC	TTTGTCAAGG	TGGGTGCCG	GGACCCCCG	24990
	GAGGAGCCCT	GCTGGACCTT	GGGAGGTGCT	GGCTGATTTG	CACCTCATGT	TGGGTGGAGG	AGGFACTCT	25060
	GGGTGGCCG	CAGGGAGTGC	AGGTGACCTT	GTCACTGTG	AGGACACACC	TGGCACCTAG	GTTGGAGGCC	25130
60	TTCAGGCTTT	CTTGCAGCAC	ATGGGGCGA	CTGTGACCC	TGACTGCCG	GGCTCTTATT	CCAAGGAGGG	25200
	GTCCCCACTG	ATTCAGATT	CGCTCAGAGA	AGGAAGGCCA	ACGGCTTAC	CACCAAGGCC	CGTGTCTT	25270
	CACCCCAAGTC	CTGAGGCAAGG	GGTCTCTGTG	CTCTGGGCTG	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25340
	GGGTGGAGG	GGGGGGGGGT	CAGAGAGAGA	GGGGGGGACA	CGGGCAGGCC	AGGGCTTGA	GGCAGAGGTG	25410
65	ATGTCGTACT	TCTCTGGTG	CCACTGTGAG	TCTCTCTGC	TCTACTCACA	CAGGTGGATG	TGACGGGCC	25480
	GTACGACACC	ATCCCCCAGG	ACAGGCTAC	GGAGGTCTAC	GGCACCATCA	TCAAACCCCA	GAACACGTAC	25550
	TGGTGTGCGT	GGTATGGCGT	GGTCCAGAAG	GGCCGCCCATG	GGCACGTCCG	CAAGGCTTTC	AAAGGCCACG	25620
	TAAGGTTCAC	GTGTGATAGT	CGTGTGCTGG	ATGTTGCTG	CTGGGATATG	AATGTGTCTA	GAATGCCACT	25690
	GTGTCGTG	TGGGTTCTG	TGGTGGAGGT	ACTTCACATG	TTTACACATG	TGTGATATG	GTGCGTTGCC	25760
70	CGTGTGCTG	GTGGTGTG	TATCTGTG	GTGCAATTT	TGTTGTTG	TGTGTTG	ACGTGTTG	25830
	CCATGGTGTG	TGTCGCTGTG	GTGTCATGT	TGTTGTTG	GTGACATG	CATGTTGATG	CTGTTGCTG	25900
	CATGTCGTG	ATGTGCTTAT	TTGTTGTTG	GTGTCATGT	TGTCGGTAC	ATATGCGTGT	CTATGGCATG	25970
	GGTGTGCTG	GCCCCCTTGGC	CTTACTCTT	CCTCCCTCAG	GTATGGTCCG	CACCATGGTC	CTCACGCTCT	26040
	GGGGTGTGCTG	TTTGGGGAGG	TCCACATCA	GGGTCTCTAC	TTCTAGCATG	GGTGGCCCTG	TCTCTGTACA	26110
	GGGGCTGGCC	TTGGAGACTG	TAAGGAGGT	TTGAGAGGA	AGTAGGGATG	CTGGTGGTAC	CTTCTGGAC	26180
	CCCTGGCACC	CCCAGGACCC	CAGTCGGCC	TATGCGGCT	CCATGAGATA	TAGGAAGGCT	GATTCAAGGCC	26250
	TGCGCTCCCC	GGACACACTC	CTCCCAAGAC	GGCCGGGGCC	CTTGGGGCTC	GGCAGGGGGT	AAAGGGGCC	26320
	TGGGCTTGGG	TTCCCAACCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTAAAGTCTG	TGCCAGGCC	26390
75	GGGTGAGAG	GTGAAGAAGT	ATCCCCTGGAG	CTTCGGTCTG	GGGAGAGGCC	CATGTGGAAA	CCCAACAAGGA	26414

Contig 2:

5 TGTGGGATG GTTTCATGT GTGGGATAGG TGGGGATCTG TGGGATTGGT TTTTATGAGT GGGGTAACAC 70
 AGAGTTCAAG CGGACCTTC TTCCCTGAGT GGCTCTGCAG GTGCTCCAC AGCTTTATTG AGGAGACCAT 140
 ATCTTCCTT GAACTATGGT CGGGTTTATA GTAACTCAGG GGTGTCGAGG CCTCCCTGG GCTCCCTGTT 210
 CTGTTCTTC CACTCTGGG TCGTGTGGT CCGTCTGTT GGTGTCGAGG CCTCCCTGG GTCAGCTCT 280
 TCCCTGTGTT CATTGGCCCTG GTGTTGGCCC TGCTCTGGT CCGTCTGTTG AATTCCTCTG CGAGTTGGAG 350
 GCTTTCTTC TTCTCTTT TTTTTTTTG TGATAACAGA GTCTCGCTCT TTGTTGCCA 420
 GGCTGGAGTG GTTGGCGTG ATCTTGCTC ACTGCAACCT GTGCTCTGGT AGTCAAGCA ATTCCTCTG 490
 CTCAGCCCTC CAAGTAGCTG GAATTATAGG CGGGCACCCAC CATCTGACT AATTGGTAA ATTGGTAG 560
 AGACGAGGTT CTCTCATGGT GGGCAGGGT GGTCTGCAGT CCGTACCTC GGTGATCTC CCACCTCGGG 630
 CTCCCAAGT GCTGGGATGA CAGGTGTGAA CGCCGCGCC CGGGCGAGAC TCGCTCTCTG CAGCTCTGG 700
 GAGATCTGCA CGGATAGCTG CTCGAGCTGAC AACTCCCTG TTCCCTCTCC AGGTCTCGCT 770
 AGGGGTTTTT CCAATTCTAGT ACTCTCTCA CAAAGAGATT TCACGTCGTC TGATTTCCCG GCTGTTCTC 840
 GCGTAATTGG TGTCTGCTG TTATCGATGG CCTCCCTCCA TTCTCTTAA GCTTGTGTTA TTGTTGTTT 910
 TCCGGCTCTC TGGGGAAA GTTGGGATTA TGATGTTGAA AACTTCTCTT CTAAACAAAG CATCTGAAGT 980
 TGCGGTTTC CCTCTAAAGG AGGGATCCCG AGGGCCCTGG CTGTTGGTGC GACCCGCTCT GGGGCTCTG 1050
 AGGAACCCGG CGCACAGCGG GAGGCTAGGT GGGGTGTTGG GAGGCGAGGT TCCCGCTGA GCCCCGCC 1120
 TCTCAGATCTC CGAGTGGCTC GGGGCTCTCA GAGGGCAGACA CACCTCTAGT AGAACTGTGC GTGAGAGGG 1190
 10 TCTAGATCTC TGCTCTCTA TGCGGATCTA ATGCGTCTG ATCTGAGGTG GAACCGTTG CTCCCAAAC 1260
 CATCCCTTC CCCACTGCTG TCCGTGGAA AAATCGTCTT CCACGAAAC AGTCCCTGGT ACCACAAATGG 1330
 TTGGGGACCC TGTGTTAAAG ACCTGCTTCGAC GAGCTCTCTG CTGCACTGTT ATATATTGGC TTGTTCTGTT 1400
 TGAGCTCAGA ATAAATTACGG ATTCTCTGTA TGCTTCTCCG CGACCTCAGA CCCATGGGCT ATTGTGGGG 1470
 GTGTTGCTG CTCCCTGGTT GGGGAGGGT CAGGGCCCAT GTACCTCTCTT GTTACTGCTT TCCAGTTGG 1540
 TTCTCAGGGT TGAATCGTAC TCGATGTTGTT TTAGGCCAT GGGGCTCTGG CCAGCTCTCTG GGGGCTCTGG 1610
 AACATGCTGA AGCACAGAGT CACCGTCCGC GCTCTCTGAT GCCTCACAAG CTGAGGGCCT CCGTGTCTCC 1680
 TGTTAGTGTG TGTCACGTCG CTGCTCACAT CCTGTCCTGG GAGGGCAGGG GCTTACAGG TCCCGTAGTA 1750
 AATGACAAAGC GTCCCTGGGG AGTCTGCAAG ATAGGAGGGT GGGGTGCGGG TCTCTCTCCTC GGGTCTCTCAG 1820
 ACTCTCTTC CGCCCTGCTG TGCGGCTCAGC CTGCTACCTCT CGACATCCCTC CAGCAGCTGG CTGGAGAGGG 1890
 20 CCGGGAGCTC GAGTGCACG TGCGCACG TGTGTTGGAT GGCAGCTGGT CACGGGGGTC TGATGTGGG 1960
 TGACTGCTGA TGGCGTTGGT TCAAGGGGGT CTGATGTTG TGACTGTTG ATGGCGGTGCG TGGGGCTCTGA 2030
 TGTTGTACTG TGCGGATGGG TGCGTGGGGT CTGATGTTG TGACTGTTG ATGGCGGTGCG TGGGGCTCTGA 2100
 TGTTGTACTG TGCGGATGGG TGCGTGGGGT CTGATGTTG TGACTGTTG GGCAGCTGGT GGGGCTCTGAT 2170
 TGTTGTACTG TGATGGCAGT CGTGGGGTCTG TGATGTTG TGACTGTTG GGCAGCTGGT GGGGCTCTGATG 2240
 TGTTGTACTG TGATGGCAGT CGTGGGGTCTG TGATGTTG TGACTGTTG GGCAGCTGGT GGGGCTCTGATG 2310
 TGTTGTACT TGCGGATGGG TGCGTGGGGT CTGATGTTG TGACTGTTG ATGGCGGTGCG TGGGGCTCTGA 2380
 TGTTGTGTA TGCGGATGGG CGGTGCTGGG TGCGTGGGGT TGACTGTTG ATGGCGGTGCG TGGGGCTCTGA 2450
 TGTTGTGTA TGCGGATGG TGATGCTGCA CAGGGCTGG TGCGTGGGGT TGCGTGGGGT TGCGGCTCTGG 2520
 25 GGTCTGATGT TGCGGATCTG TGCGGATGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 2590
 GTCGTGGGGTG TGCGGATGTTG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 2660
 GATGGCAGTCG GTCGACAGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 2730
 GACTGTGGAT GGCAGCTGGT GGGGCTGATG TGCGGATCTG TGCGGATGGG TGCGTGGGG TGCGTGGGG 2800
 GACTGTGGAT GGCAGCTGGT GGGGCTGATG TGCGGATCTG TGCGGATGGG TGCGTGGGG TGCGTGGGG 2870
 30 GACTGTGGAT GGCAGCTGGT CCGGGGGGTC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG 2940
 GTGGTACTG TGCGGATGGC TGCGGATGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3010
 GTGTTGGTACG TGCGGATGGC TGCGGATGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3080
 ATGTTGCTG TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3150
 ATGTTGCTG TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3220
 GGGTCTGATG TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3290
 GGGTCTGATG TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3360
 GTCACAGGGG TCTGATGTTG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3430
 GCGGCTGTTG GCGGCTGATG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3500
 TGCGGCTGATG TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3570
 35 GCGAGTGGAG TCCCAGGTG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3640
 GAAGCTCTCC AGGCCTCTC TGCGGATCTC CCGGGCATCTC GGCGTGGGG CGAGCTCCAC CGTCCCTGATC 3710
 GGAAGAAACA ATGCGCAGC TCTGGGGGG GCAGGGCACA TTGCGTGGCT ATGCGCTCTC CTGCGGCGC 3780
 AGGTCTCTAC CTTGACAGAC CTCGACGGCT ACATGGCACA TTGCGTGGCT CACCTGCAAGG AGACCAACCC 3850
 GCGTGGGGTG CGCCGCTGCTA TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGGATGGC 3920
 40 GTGGGCTCTC CCGGGGGCAA TCACTGGGGT CATGACGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 3990
 GAATGAGCTG TGATGGGGG ATGATGAGCT TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 4060
 CGCAGCTCCC TGCTTACCGC CACCTGCTCA CGTGGTACTG CGCGCCCTCT CGCAGCTCTC GCGAGTGGCTT 4130
 TGTCATGAT TGCGTAAATG TCTCTCTG TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC 4200
 TTTAGGAGGG CAGGGCATGT TGCGGATGGC TGCGTGGGG TGCGTGGGG TGCGGATGGC 4270
 45 AGGAAACCTG TCCCCCTCTG TAGGAGGACG TGCGTGGGG TGCGGATGGC TGCGTGGGG TGCGGATGGC 4340
 TGCGGGTCT TGCCACGGT CGCCGCTGGC TTGGCGAGAT TGCGTGGGG TGCGGATGGC TGCGGATGGC 4410
 TGCGAGATGCC TGCGTGGGG TGCGTGGGG TGCGTGGGG TGCGGATGGC TGCGGATGGC TGCGGATGGC 4480
 CTGGGGCAAT TTCCCTGGCT CGCCAGGGTGG GGGTGGAGGT GGCGCTGGGG TGCGGATGGC AGACCCCTGTG 4550
 CCCGGCAGCT GGGCAGGAC CTCGATGCA TGCGTGGGG TGCGGATGGC TGCGGATGGC TGCGGATGGC 4620
 CCCAGCTGGA CCCACAGGG TGCGGATGGC TGCGTGGGG TGCGGATGGC TGCGGATGGC TGCGGATGGC 4690
 CTGCGAGAC CGCCGCTGGC CGCCGCTGGC TTGGCGAGAT TGCGTGGGG TGCGGATGGC TGCGGATGGC 4760
 50 CACTGACCCC AAAAGGGGAGC GAGGGCTGG TGCGGATGGC TGCGGATGGC TGCGGATGGC TGCGGATGGC 4830
 CCCATGTGTC TCCCGCTGCG TTGCGAGAG TGCGGATGGC TGCGGATGGC TGCGGATGGC TGCGGATGGC 4900
 TTCTTACGGT TCATGTCGCA CGACGGGGTG CGCATCAGGG TGCGGATGGC TGCGGATGGC TGCGGATGGC 4970
 CCCCTGGGGT GGCGTGGGG TGCGGATGGC TGCGGATGGC TGCGGATGGC TGCGGATGGC TGCGGATGGC 5040

	CTGCCCCGGGG	CCACCAAGAGT	CTCCTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTG	GCTTGGCGCA	GCCGGAGCGG	ACGAGGTGCC	ACACGAGGCC	TGGAATGCC	5180
5	AAGCCGGGTG	TGGAGTTGCT	CCTGGTGGG	GGACGAGGGG	CGGGGGGTG	GTCCTGGCTA	GGTGTGCGCC	5250
	GAGCGTTG	GCTGAGCTC	TGTCAGCTCC	AAGTTACTAC	TGACGCTGG	CACCCGGCTC	TCACACGCTT	5320
	GTATCTCT	CTCCCATA	AAAAGGATT	TATCCGATT	TGATTCCTGT	CCTCTGCTG	TGACCCCCCG	5390
	GAGGGCGGG	GCTCTTCTCT	CTGTGACTAG	ATTCCCATC	TGAAAGTGC	GGGGTTGAC	GTGTAGTTG	5460
	CTCTCTCGG	GGGGCCATG	GGCAGGCTGG	CTGGGCTACCA	CAGCCACTG	GGGGAGGAGC	5530	
	GTGACCCACA	CTCACGGTGG	TAGAGCCACA	CTGCTGGTG	CCACATCAGC	TCTCTGGAT	TTAAAGTAAA	5600
10	ACACACACCC	TCCCAGCAGG	CATCTGCCG	GGACCCGTG	TGTGCTGGG	GAGAGTGGTA	GCACCGAGGA	5670
	AATTCTGCA	CACTCAAGGT	CATCAGCAAG	GTCACTCCG	GTCAGGTGG	ACCTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTC	AGCCGATAAA	GGACTGTGCA	CAGCTCGGA	ACCTTTTATT	TAAAATATA	ACTATAATT	5810
	ATTGCAATT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGG	AAAACACAAA	TTGCACATGG	CAGCAGAGT	AGAGTTGGCC	GAGGGACACG	5950
15	TGTGACAGG	GGCCCCAGGC	CGACAGAATT	CGCTGACAAA	GTACACCTCC	CAGAGAACCC	6020	
	ACCAGGGCC	TCTTCGTGG	TCGTGAATT	TATTAAGATG	GATCAAGTC	CGTACCGTC	ACGTGTGCCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGTGACTCTG	GGCTCCATCG	CTCGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGCTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCC	ATCGTGTG	AGTTTGGCT	CTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAACT	AGAACACAGAG	AGAGTTCCC	ATCCCAGTG	CTCACAGGG	6300
20	CGTATCTGCT	TGCGTTGACT	GGGGGCTGCT	GGGGGACTGC	TAGAGTTGG	GCTGTTGCTT	CTGTGCAAAA	6370
	AGTCAGTCC	TCTGGCCAT	CACTGTATA	TCTGCACAGG	CAAGGAAAGC	TCTTTCTT	TTCTTCTTT	6440
	TTTTTTTTT	GAGACGGAAC	GTCACTGTT	TCTGCTGGG	CTTGAGTGC	GTGGCGGAT	CTCAACTCAC	6510
	TGCAACCTCC	GGCTCCGGG	TTCCAGCATT	TCTCTCTAC	CAGCAGCTCCC	ACAGAGCTGAG	ATTACAGGCA	6580
	CCCACCCCTC	GGGGCTGGCT	AATTTTGTA	TTTTTAGTAG	AGAGGGTTT	TTGGCATGTT	GGCCAGGCTG	6650
25	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCACGCC	CAAGCGGAAAG	GGCTCTTTT	AAGGTGACCA	CTTATAGGC	TTCCCGGAAA	TAACAGGTC	6790
	TGTTTTTGC	TGAGGTGCA	AGGGCTCTT	AGGCAACAGGA	GTGGCGCTCT	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCCTCTGTG	GCACCTTTAG	GTCCACGGG	GCTATTCTGC	TCTCACTGTT	6930
	TGCTGTA	GGCAGGCTTG	GCATCTCTG	TTGGAGAGTT	CTGCTCTCTC	GTGGGTGATG	CTGAAACTAG	7000
30	GGGAAGGTT	GTATCGTTG	GGCCGCAGCG	GTCATCATGA	GGGTGATGAG	TCTTCACGG	TGGACAATT	7070
	CCTTGGAAAA	AAAAAAAGGA	GTCCGGTTAA	GCATTCTATC	CGGGTCAAGT	GTCTGGTCT	GTGAATAAAC	7140
	TCTAAGATT	AAGAAAACCTT	ATGAAAGAA	AACCTTGATG	ATTCAGAGCA	AGGATGTGGT	CACACCTGTG	7210
	GCTGATCTG	TTTCAGCGC	CCCAAGCTAT	GGTGGAGTGG	GGGAGCAGGG	ATGTTTGTG	CAGAGSTCTC	7280
35	ATCTGGTATG	TTTCAGGGT	GTTGGCCG	TGAATGGTAG	ACGTGCGTT	TGTTGTGATG	AGGTTCTGTTG	7350
	TCTGTTGTTG	GTCCTGGTTG	AGTGTACGCA	GTGTCAGGAC	ATGCCCTGCC	CGTCTCTCAC	CTGTTGCTTC	7420
	CCGGCCCTAG	TCTCTACCTC	GGTCCCTG	AGTCCCGAG	GCTCTCATC	TCTCTCACGCT	GCTCTGCAGC	7490
	CTGTGCTACG	GCGACATGG	GAACAAGCTG	TTTGGGGGA	TCGGGGGAA	CGGGTGGAGG	CTCTCTCTCC	7560
	CAAGGGGGC	TTGGGGGGG	TTGGATTTG	TTTGGATGCA	TCAGCTGTTA	ATATTCTGG	TGCTCTGGAG	7630
40	ACCATGACTG	TCTCTGTTG	AGGAAACAGGA	CAAGGCTCA	GGCCCTCTT	GTGATGAGG	7700	
	GGGTTGACA	GCCTGAGGAC	TGCGGGCTCC	ACGAGGCTC	TGCCAGCGG	CCATGTCAG	AGGCTCAGG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCTCTGC	ATGATGACCA	TGTAATTC	ACACCCGAGGA	AGCACACCG	7840
	CTTCTGTCAC	TCTCACCGAG	TTCCGTTAGG	GGCTCTGGGG	GGATGGGCT	GGTGCAACCT	GAGGCCAAC	7910
	ATCTCCCAGC	AGGCCCTCGA	CAGGTGGCC	GGACTGGGGC	CCTCTTCA	CCATTGCCC	TCCCACCTGC	7980
	ATGGGGTCA	CAACCAAGGA	TAATATCTG	GGCACACAC	GGCAACCTA	TGTTGTTCAA	CTCAGCTG	8050
45	TTTTATTGAC	AGCAGTACT	TTTTTTTT	TAATACTTA	ATTTCTAGG	TACATGTCA	CGACGTCAG	8120
	GTAGTGTACA	TATGTATACA	TGTCCTATG	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCCCT	TGCTTACCTC	CCCCACCTCC	CCCCATCCAT	GGAGCCCT	GGTGTGATG	TTTCCCCAAC	8260
	CTGTGTCCTA	GTGTTCTCAT	TGTCAGTC	GGCCTCTGTA	GTGAGAACAT	GTGGTGTG	TTTTCTTTC	8330
	CTTGAATAG	TTTGCTCAGA	GTGATGGTT	CCAGCTTCG	CCATGTCCT	ACAAAGGACA	TGAACATCATC	8400
50	CTTATTTAG	ACTGCTAGT	ATTCCTGTTG	GTATATGTG	CCACATTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGGATTGG	TTGGGTGCA	AGTCTTCT	ACTGTGATA	GTGCGGCAAT	AAACATACGT	GTGATGTG	8540
	CTTATAGCA	GTATGATT	TAATCTTGTG	GGTATATACC	CAGTAATGGG	ATGGCTGGT	CAAATGGTAT	8610
	TTCTAGTTG	AGATCTCTG	GGATATCACA	CACTGCTCTC	CACAACTGGG	GAATAGTTT	ACACTCCAC	8680
	CAACAGTGT	AAAGTGTCT	GGTGTCTGGAG	AGGTGTGGA	CAGCAGTTAT	TTTTTATGTA	AAATAGTATC	8750
55	ACTGAACAA	CAGACAGTTA	GTGAGGATG	CGTCAGGAAG	CCTGAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGT	TTTCTCTGTC	CATCTTTGA	AACCTCTAGT	CCAACTTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTCTTCTAT	TAAGGTCAA	TTCTCTGATT	GGAAATAAGT	TATGTAACAG	AAACAAAAT	TTCTGTACA	8960
	CACAACTTGC	TCTGGATT	GGAGGAAAGT	GTCTCTGAGC	TGGCGGACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCCATGGT	CATGGGGG	GGGGCTTGGG	CTGGTGGGG	ACACAGTGC	CCATGCCCG	9100
60	CTTCTCTGG	ATAGGAGCTG	GGTCTCGGAT	CATGTCAGG	ACACAGCTG	CCATGCTGGT	AAAGGGCAC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTCTCCAGC	CCCAGCTTTC	TTACCGTCTT	CAGTTATTT	TCCCTAAGAG	9240
	TCTGAGAAGT	GGGGGGGGC	CTGATGGCC	TGCTCTGCTG	TCAGCTGGCA	CAGAAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGATGAGG	AATTCTGTA	CGATGTAAC	TGAGAGRC	CGTCTGTTG	9380
	AAAGAATTT	AAAGTTTCA	TTAACGGCT	TTGGAGAATG	TTACTTTAT	TATGGCTG	AAATTGTTT	9450
65	GACATTCAGT	CCCTCTGAGA	CAGATTAAC	GTAAAGTGC	TTAGAGTAC	CTTGTGTTG	ATTTCCCTT	9520
	ATTTTAGGT	GCTCTCTGGT	TTGGTGGATG	ATTTCTGTT	GGTGACACCT	CACTTCACCC	ACGGAAAAC	9590
	CTTCCTCAGG	TGAGGCCCCGT	GGCGTGTGTC	TGTTGGGACC	TCCACAGCT	GTGGGCTTG	CAGTTGAGG	9660
	CCCCCTGTC	TGCCCCCTGGG	ACCCAGCGGT	TGCTCTGTC	AACTCTCTC	TCTCTGCCCC	TGCTGGATCC	9730
	GCAAGAGCG	AGGGCGCTGG	CGTGCACCC	AGGCCCTGGG	GGCGAGGGG	ACCTTCGGGA	GGGAGTGGGT	9800
70	ACCGTGCAGG	CCCTGGCTCT	GCAGAGACGC	ACCCAGGTGA	ACACGTGGT	GAGTCAGGC	GGTACCTGG	9870
	CTCCCTGCTG	TCTTGGAAA	GTCAAGAGT	GGGGCTCTG	GGGGCCCCAGT	GAGACCCCCA	GGAGCTGTC	9940
	ACAGGGCTG	CAAGGGCCAG	GGGGCTGGCT	CTCTCCCAGG	GTGACCTG	GGCTGCGGAG	AGCAGGAGGT	10010
	GCTGAGTGA	CTGGCCACAA	GGGTTCTGCTG	CGGTACCGTT	CCTGCGTGGG	GTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGT	CTGCTGTTCT	GAACCAACGGA	GTGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTGAATC	AAACTAAAT	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTC	ATGTGTC	10220
	CCTCAAGGGC	GCCCCACAGA	GGCCGGGGC	TTGTTTAAA	GTGCAATTG	ACGAGGGAC	AGAAACCTG	10290
	AAAGCTGTA	AGGGAACCC	CAGAAAATGT	GGCCGCCAGG	GGTGGTTCA	GGTGTGTTG	TGGGCTGTC	10360
	TTGTGAAAC	CCATTGGAC	CGGCCCTCCA	AGTCCACCC	CCAGGTC	CCTCCAGGGC	CGCCCTGGG	10430

5	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCG	AGCCCCGAGC	ACAGCAGGCT	GTGACACATT	AAATCCACTA	10500
	AGATTCACT	GGGGGGAGCC	CAGGTCCTAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGGCTG	CTTCCTGTG	GCAACTTCTG	GAGGGTGC	GGCAGGGAGG	TGGCTCAGAG	10640
	TGTATGTTGG	GGTCCACCGG	GGGGCAGAAC	TCTGTCTG	ATGAGTCGGC	AGCCATGTA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGGAA	TGACCCAGGG	GAGCTGCGC	GCTGGCGAG	GTCCCAGGG	CAGGCCACAG	10780
	GAAGGGCGAG	GGGACCCCGG	GGGCCACAGC	AGAGCCCGCA	GGAAAGGGAAG	GGGATGCCA	GGCCAGAGGC	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGTGA	CGAGGCTCAT	GACTCGGG	GGGAACCTC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCTG	AGCAGGAAC	10990
	CAGAACCCCT	CCCTTGTCT	AAAGCACACG	AGATGCCCTC	AGGGCATCTG	GGAGAAAACA	GGCAAAGTCTG	11060
10	TTGAGAAACG	CTTAAAGAAG	AGGTGGGATG	GTGGCAATT	CTTGTCCAGA	TTTAGTCTG	CCCCGGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGAA	CACATGAGAT	GGACCATCAC	AGAGGCTACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCCTGGCT	GTCCCTGGCT	CAGGCCAGGT	TCTTGTGTC	TCACCTTACCT	11270
	GTCTGCCCCG	GGAGAACAGGG	AAAGCACCCC	GAAGTCTGGA	TGAGGCTGG	GGCCAGGCTG	CTCAGAGCTC	11340
	CTGCCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTTCTC	GGGGTTTCC	AAAGCATTCA	ACAAGGGTGT	11410
	CAGGTTACCT	CTCTGGTGAC	GGGGGGGACAT	CTCTGGGCTG	ACATGGCC	TCTGCC	GACCCCTGGTC	11480
	CGAGGTGTC	CTGAGTAGTG	CTGGGTGTC	AACTTGCGGA	AGACAGTGGT	AACTTCTCCT	GTAGAAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCGGGCCCA	CGGCCATT	CCCTGGTGG	GCCTGCTGT	11620
	GGATACCCCGG	ACCTCTGGAG	TGCGAGGCC	CTACTCCAGG	TGAGGCGACC	TGGCCGGAAAG	TGGAGCCCTG	11690
	GGCCGGCTGG	GGCAGGTGCT	GTCGCGGCC	CGTGTGCCCC	ACCTCTGCTT	CCCTGGTGG	CAGGGCAGTC	11760
20	CCAATCCCAA	AGGGTCAGAG	GGCACAGGGT	GGCCCTGTC	CCATCTGGG	CTGAGCAGAA	ATGCATCTT	11830
	CTGTGGGAT	GAGGGTGAC	ACAAAGGGG	CAGTTTTCTG	GTCTATT	GTAAAAGGAA	ATGGTGCACC	11900
	AGACCTGGGT	CACTGAGGT	GTCTCAAGA	AGACGTCGG	ATGCTTCC	GGCCCTGCTG	11970	
	GGCGTGA	TCTCAAAACC	GAACACAGGG	GGCCCTGCTG	GCATGAGTC	TCTGAA	GAGACCCCTG	12040
	GGCCGTGCTG	GGCGTGA	TCTCCGA	GAAGACAGTC	AGGAGCTT	GGGCCCTT	TGGCGTGTG	12110
	GTGAGCCCCA	CACTCAAAAG	CTCATCACAA	GTCTCACAGG	TGCCATGAGT	TCTGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGG	TCTCTACAAA	ATTCTGGG	CTTGT	CAGACCCCGA	12250
	GAGGTCAGG	CCCCGCTCA	GGGTCAGACA	CAAATGTA	GAAGATGGA	ACAGATGCA	AAATCTGTG	12320
	TGTTCTTTT	ATGAATAAA	AGTATCAC	TTCCAGGCG	GGCAAGG	CTCAGACCA	TAATCCCAG	12390
30	ACTTTGGGAG	GGCGAGGTG	GTGGATCA	TGAGGCGAGG	AGTTT	GAGGAGG	AAACATAGTGA	12460
	AATTCCATT	CTACTTAA	AAATCAAAA	TTAGCCTGCT	CTGGTGGC	ACGGCCTG	TCCCGCTAT	12530
	GGGGGAGGCT	GAGGCAAGG	AATCTTGA	ACCCAGGAGG	CAGGGTCA	AGTGGCCG	GATCACACCA	12600
	CTGCACTCCA	GGCTGGGCA	CAGAGTGA	CTTCATCTT	AAAAAAA	AAAAAGTATC	AGCATCTCA	12670
	AACCATAGTG	GACAGGTGTT	TTTTTATCT	GTCTCTG	AAATTTACT	GTGCTG	TAGAGGCCG	12740
	AACTGGGGT	GGCTTCTCT	GAAGGACACA	CCTTCATGGG	AGAGAAA	ATGTTG	GGTTGTTAA	12810
	CCAGAGGT	AAACTGGGGT	CCTGCTTC	TGAGTAA	GCATGAGTC	GGACTT	TCTTCCAGA	12880
	ATGCTCTCTG	GGGTTGCTT	CATGGGGGAG	CAGCAGGT	GGACACCC	GTGATGGGG	AGCAGCAGG	12950
	GCAGACGCC	TCATGATGGG	GGAGTGGCG	TGCGACAC	CTTGTGCT	GGTCCCAGC	ATGCTCTGT	13020
	TGCA	GGGGCTCCC	TCCTCCAAAG	GATGCCG	CTCTGCTC	CCCAGTC	CTGCTTCT	13090
40	TACCTGGTCC	GGGCTTCCAC	TGGCTTGTG	TGCTGATG	CCACATT	TGGGCTCC	GCACCTCTC	13160
	GCCTCTCCA	GGCACCTCTG	CAGTGTGTC	CATACAGTC	AGCTG	TGTC	TTATTTGCT	13230
	CCCCATGAA	TGTTTTTTT	AGGACAGGG	CCCCTGGT	CAGCCTG	CAACGATCA	GTGATGTTA	13300
	TTGAGGACAA	AGGGACAGAC	AAACAAATCGA	GGAAAATGG	TCTCTCTAA	ACACATGCA	AAGCCACAGA	13370
	GGCTAGTGA	GGATGGGTG	GCATCAAGTC	ATCAGATG	GGTCA	GGTCA	CAGAATATTC	13440
	AAGGCCACTT	GGTCAGAGTG	TGTGCTGCA	GAGGGGCTC	TTAAAGCTC	CGAGTGGGG	CAGTGGT	13510
	CCATACCTAG	GGTGAACCTA	CATCCTCTG	TGCTGAAGT	TACAGCAGA	GCTTGAAGG	CATCTGGGAG	13580
	AAGAAAAC	CAAAATGAT	TAAGAAAAG	GGAAAAGGA	AAAGTGGTA	ATGGAATTT	TCTTGTCC	13650
	ATTTTAGTCT	CCAAAACAC	AGCTCAGATG	GTAGAATG	GTCAAGACT	GTGAGCAGAA	CAATAGAAC	13720
	AAACCGAAGC	CTTATCTCTG	AAACACCTG	GTAAATGTTG	TGATG	GGC	AAAGAGACTG	13790
	TGTGTGAA	TTTTTTCT	GAGAAAAGCA	ACTGGAAGCA	AATAAGTGT	GTCTT	TACAG	13860
	AGCAGATTCT	AGGTAGAGA	GGAGACACAT	GCACAAACA	CAAGAAC	AAATAAAC	AAAGACTCAA	13930
	AGGGAAAGGG	GGTGAACGTT	CCCTGGGTTG	GTGTGGGG	AGGACACACA	GGGAGGGG	TGAAACCGT	14000
	GAGGCAACGG	GCATTGCTT	CACTGAGAG	AAACTCAGG	TGCTGAGG	ACAGT	TGGGCTTCC	14070
	CTGGAGCGTT	TGTCAGCTG	TTTTTATTA	GGGCCCTGT	GAGGCTCTG	CAATT	TCTCACTT	14140
	TTCTCCTAAC	CACTCTGAGAG	GTAGAGGGAG	AAAGGCTTCA	GGGGAGCAGC	CGGCCCTG	CACCCAGCTG	14210
	GCACAACTCA	TCCATGAGG	CAGCAGGGG	TCTCTGCTC	GGGGCTCTG	TCTGCCC	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCG	GAGGCCAAGG	TCTGTGTTG	GATGGCTG	AAAGAGGAA	14350
	TGGAGCTTG	ATTCAGACTT	TGCAAGGAG	TGCAAAAGAAA	TGCATG	ACTGACACGG	14420	
	AGACCCATCC	CTCAAAAGAAA	GGCAGCTGAA	AGCTG	GGACAGTC	CCATCCCTCA	TGCTGGCTC	14490
60	TTTCTGGG	TTGCGAAGAG	CCAGCTAG	GTTGAGGCA	GCTTGAAGA	GGTTC	TGTTCTG	14560
	TTTCTGATG	AGTCTCTCA	ATGCTCTG	TCTTCCCAGT	AATTCTACTT	CTGAAGT	CAGACATTAT	14630
	CACGGGTCTT	ATTATCATT	TCCAGTGTG	CAGGGAGGGG	GACTG	CCAC	AAACAGTCAC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCA	ACAAAAC	CTCTGCTT	GGGAGG	GAACCTGCCC	14770
	TTGAAGATG	TTAAATGGCA	CAAAACGTTT	TTTCAATG	AGCAGTGT	AAAGCTG	GTAAAAGAAC	14840
	ACACCCCAAGG	AGCCTGCGT	GAATGTCATG	TGTTGTCAT	TTGGACATG	GACATCATG	GGCAGTGT	14910
	GGTGGTGGAG	CCCTGGGAGG	CATCGGTTGG	ATGCTCTCAT	CCTGAGGCTC	TGAGGACAC	ATGTTGCCC	14980
	CGTGGCACTA	CTGGAGGCTC	GTGGAGGCTG	TGCGACCTCTG	CTCTGCTC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCTCA	CGCCCAACTC	AGTGTCTC	CACAAAAC	CTGAGTCAC	CCTGAGTTC	CTCGAGGGAC	15120
	GCCCCGGAGC	CAGGGCTC	CAGTTTATTA	TGTGTTTGT	GCTGAGTT	GTGCA	CATCAGGGCA	15190
	GATGATGAGT	GCACAAACAC	GGCCGTGCGG	GGTTGGGATA	CACTCAACAT	CAGTAGCCAG	GTCTGGT	15260
	AGTTTGGTCA	TGCAAGACTC	GGATGCGATG	TAGCATTTG	AGTCCATG	GTGAGCACC	AGCCCCCTCG	15330
	GGCTCAGCG	CATGCCCCAG	GCAGGACAA	GAAGCAGGGG	GAAGGCAAGG	GGCTCTT	AGCAAGCTT	15400
	GCAGGAGGGG	GGCTGGGTTG	GGCAGGGCAG	CTGAGTCTGA	CATTCCCC	TGTTGTC	CTATGCCC	15470
	ACCTCCATCA	GACCCACTC	CACCTTCAAC	CGCGCTTCA	AGGCTGGGAG	GAACATGCG	CGCAAACTC	15540
	TTGGGTCTT	GGGGCTGAAG	TGTCA	GGAGGCTG	TTTGTG	GGAGGCTG	TGGTCA	15610
	AGAGTTCAGA	GTCAGGAGG	TGTTGCGCA	AGTATGTTG	TGTTGCGT	GGCGCGT	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACTA	AGAGTGCAC	TGTA	TGACCTG	CACATG	TGTGATG	15750
	TGACATGAA	GGCATGGCG	TGTGCA	GGTGTGCAAG	GGCACAAGTG	TGTGCA	CGAATGCA	15820

5	CCTGACATGC ATGTTGTTG GTGCCAGTC GTGTGGGCAT TCACGTGAGG TGCATGGGT 15890 GTGTGAGTAG CATGTGCA CATAACATGT ATTGAGGGGT CTCGTGTC ACCCGCTAG GTCCCTCAGCA 15960 CCAGTGGCAC TCTTACAGG ATGAGACGGG GTCCCAGGCC TTGGTGGGC GAGGCTCTGA AGCTGCAGCC 16030 CTGAGGGCAT TGTCCCATCT GGGCATCCGC TCTCTGTG GCTTCTGTG CCACTCCCC 16100 TCTCTGTG GCTTACAT CCACTCACT CCCCCTCTG TGTCGACTCC CGCGTCCACT CCCCCTCTG 16170 GTGGGACATCT GGGTACACCT CCCCTCTG TGTCGACTCC CGCGTCCACT CCCCCTCTG 16240 CTTGGGGAG CTCGGGGC AGGCAAGATGA CACAGAGTGA AGGAGGATAG TGTCGACTCC CGCGTCCACT 16310 GGTAGGGCC AGGGGGGATT TCACTGGGA GAGGAGTGA TTCTGTCAA AATGTTCTTC TTTCTGTG 16380 CATCTGAATG GATGATAAAG CAAAAAGTAA AAACCTAAAA TCCCAGAGAC GTTCTACCG TTTCTACTC 16450 TTCTGGGG ACTCTAGGT AACACGGTCA AGACGGTGTG CACCAACATC TACAAGATTC TCCCTGTGCA 16520 GGCTACAGG TGAGCGGCCA CCAAGGGGTG CAGGGCCAGC CTCCGGGAC CTCCTGGGCT CTGTCACCT 16590 CTGACCCGGG GCTTCACCTT GGAACCTCTG GGTGTTAGGG GCAAGGAATG TCTTACGTT TCAGTGGTGC 16660 TGCTGCCGTG GCACAGTCT TGTCGCTGCA CTGGACCTGT CTCCATCTCT GGGTAGTGTG 16730 AGGAGCCGGT GTGGCCCGAG TGTCGCTG CACTGGGGT GGGACGTCAT GGAGGCACTC 16800 CCAGGGCAGC AGGGGCATGG GTAAAGAGA TGTTATGGG GAGTCTTACG AGAGGAGGCT GGGAGGGT 16870 CTGAACAGTA GATGGAGAGT CAGATGCCG GAGGGTTGG GGTCTCAGCA AAGGGGCCG AGGTGGGTG 16940 AGGTGAGGGT CGCTGGCCCC ACCCCCCGGG AGGTGAGCA GAGCTGTCG TCCCCACACA GCCCCGGCAG 17010 CACCTGTGCT CTGGGCATGG CTGTCCTCT GGAACGTTCC CTGTCCTGG TGTCAGGGG GTGCCCCCTG 17080 CAAGAGATGA CAACCTTATC ACAGAGGGGA GGGCAACAT TGTCGCTGCA CAGGGCCAGC TTCTGCCCTG 17150 AGTCAGGGCA GTGGGGTGGCA CAAGCTCGG TGTCGCTGCA AGGGCAGTC GGGCACACACA GGGCCGGGCC 17220 TCCACCTCAA CAGGCCCTCC GAGCCACTGG GAGCTGAATG CCAGGAGGCC GAAAGCCCTCG CCCCCTGAGG 17290 GCTGAGAAGG AGTGTGAGCA TTGTTGTTAC CCAGGCGGCA GCTGGCCGCA ATTACCTGTG ACACTTGATG 17360 TGAAATGAGG TCGTCGCTA TCGTGGAAAC CAGCAAGGGG CTACCGGGAG ATGTTTCCAT TACAAGGTG 17430 TACCATGAAA ATGGTTTTA ACCCGAGTGC TTGCGCTTC ATGCTCTGGC AGGGAGGGCA GAGGCCACAGC 17500 TGCATGTTG CGCCCTTGGCA CCAAGCTCCG AGGCTGGGA CCAGGCTGTC TGAGTCCAG GGTGGCTCG 17570 GCTCAGACCG CCCCCTCTC TGCTCTCTC TGCTGCTCA ATACCTCCCT CGTGTGATC TCCCCTGACCC 17640 GTGCGTGGGC CTCGTGCAA GCTGTTGAC TCTTCTCGG AAACCTTGG GGTGTGCTGG ATACAGGTG 17710 CACTGAGGAC TGAGGGTGTG TGACACTTG GTTGGACCCCA GGGCTCAGCT GGGCTGCTTG GGGCCCTCTT 17780 GGGCATGAG GAGTTCCTC GAGTGGAAAAC TCTCTGGAAA CTCCCGGGC CATGTGACCT 17850 GCCACCTGCT CCTCCCATAT TCAGCTCAGT CTGTCCTCA TTCCCGGAC AGGGTCTCTA GCTCCGAGGA 17920 GCTCCCGTAG AGGGCCCTGG CTCAAGGGCAG GGGCTGCTGAG TTCCCGGAC CATGTGGGGA CCCTGGGTA 17990 GTCGCTTGTG TGTTGAGGCC TGAGGAGGG AGGATGGCAT GGGCCAGGG CGCTTCCAA ACACAGAGTC 18060 AGGCACGTGG AAGGCCAGG AATCCCTTC CTCGAGGCA GGAGTGGGAG AACGGAGAGC TGGGCCCCGA 18130 TTTACGCGCA GCGGAGCTGC AGTGGGGAG GCTGTGGTGG TCCACCTGGC GCTGGGGCG GGGTGTGATT 18200 CAAATCCGCT GGGGCTGGC TCTCTGGCC CGTGTGGCC GGGCCCTCAC AGGGCTGGG GGTGGACGCC 18270 CCGACCTCTA CGAGGTGGCT ATTCTCCCT TTGGAAGAGA GCCCGCTCAC CATGCTAGGT GTTTCCCTCC 18340 TGGGTGAGGA CGGTGGGGCT GTGGAACACC CGGGGACCTTA GGCTTATTTA TTGTTTAAA ACATCTCGG 18410 CCCTGGCTTC CGTTGTTGCT AATGGGGAA AGAACATCCC ACCTCAGCAG AGTACTGAG AGGTGAAAC 18480 CGGGGTGCTG GTTGACTGG TGTGATCTCA GGTCATTCCA GAAGTGGCTC AGGAAGTCAG TGAGACCAAGG 18550 TACATGGGGG GCTCAGGCTG TGGGTGAGAT GAGGTACAGC GGGGGCTGAG CAGCTGGGAGT 18620 CATGGGGGGC CTAGGCACTG GGTGAGATGA GGTCACGGG GGGCTCAGGG AGAGGGTCAG ACCAGGTACA 18690 CGGGGGCTCT GATCACACGC ACATATGAGC ACATGTGCA ATGTCGTT TCATGGTAGC CAGGTCTGTG 18760 CACACCTGCC CAAAGGGTCA AGGAGGCTGA GAGGGGAAAG ATGGAGGCTG AGGGGGTGG CCGGGGGCT 18830 CACACCTGTA GTCCCAGCAC TTGGGAGGG CGAGGAGGA GGTACATCGA AGGGGGTGG CCGGGGGCT 18900 GCCTGAGCAA CATAGTAGAA CCCATCTCT ATGAAAATA AAAACAAAAA TTAGCTGAAC ATGGTGGT 18970 GGGGGGCTG TTCCAAACT TGGGAGGGAG AAGTGGGAGG ATCACCTGGC CCCATGGGAGT GGAAGCTGCA 19040 GTGAGCTGAG ATTGCAACAC TGTACTCGAG CCTGGGTGAG AGAGTGGAGG CCCATCTCAA CAACAAACAA 19110 GAAGACTGAC AAATGCAAGT TCTGGAAAG AAACATTAG TAGGAACCTA ACCTACACAC AGAAGCCAAAG 19180 TCGGTGTCTC GTGGTCACTG AGATGAGATG ATGGGCTCTC ACACCATCAC CCCAGACCCA GGGTTTATGC 19250 50 ACCACAGGGG CGGGGCTGCA AGAAGGGATG CGGGGACCTG TGATATACGA TGACATCAAG GTGTCCTGAC 19320 GAAGGGCAGG ATTCACTGATA AGTACCTGCT GGTACACAAG AGAACATGGA TAAACTGGAA ACCTTAGAGG 19390 CCTTCCCGGA ACAGGGCTCA ATCAAGACGG AGCATGGGG CCTGGCATCC AGGATGGAGC TGCTTCAGCC 19460 TCCACATGGC TGTTCATACA GATGGTGCAC AAACAGCGAG TGACCTGTG CACACAGACA CACCGAGCTA 19530 CTCGCACACA CAAGCACACA CACAGACATG CATGCATGCA TCCGTGTG TGACACCTGTG CCCATGAGGA 19600 AACCCATGCA TGTCGATCCA TGCAAGCCAGA CAGGGCACCCG TGGGGCCATCC CCCACACCCA CGAGCACCCG 19670 CTGATTAGGA CGGCTTCTC CTGAGCTGT CGGCCATCTT CGGGGTTTC ACAGCATGTCG GCTGAGCTC 19740 CCATTTCATC AGCAAGTTG GAAGAACCC ACATTTCC TGGGGCTCAT CTCTGACAGC GCGCTCCCT 19810 GCTACTCCCT CCGTAAAGGC AGAACAGCG AGCATGGGG CCTGGCATCC AGGATGGAGC TGCTTCAGCC 19880 CTGCTGGTGT TAGTGTGTC GAGAGCTGAG TGAATCTGGG CTAGGAGAAT TCTTACCCCT TTTCGCACTA 19950 60 GGAAGTGGTT TAACCCAAAC ACTGTCAGGC TCGTCTGCC GCCCTCTCGT GGGGTGAGCA GAGCACCTGA 20020 TGGAAGGGAC AGGAGCTGTC TGGGGCTGCTC CATCCCTCCC ACCTGCTGTG CGCTGGGGAA CGGGTGGGG 20090 GCCCTGGCTC TCCGTGTTG CCCATGGG GATTGGGGG GCGCTGGCTC TCTGTGTTG CCGGGGGCT 20160 GATTGGGCTG TCTCCCGTCC ATGGCACTTA GGGCCCTTGT GCAAAACCG GCGAAGGGCT TAGGAGGAGG 20230 CCAGGGCCAG GCTACCCACCA CCCCCTCAGG AGCAGAGGGC GCGTATGCC AGCACAGAGC CCCGGGGCCG 20300 CCTCTGCTC CTGAGCTCCG CCTCTCTGCC AGCAGAGGGC GCGTATGCC AGCACAGAGC CCCGGGGCCG 20370 CTGAATTCA AGCCATGTG AGACCTGCGT CCTGAGCTTA ACAGCTCTCA CTTCTGTGTT TTCTGTGTT 20440 GTGGAATTTC CACCTGGAGA AGCCGAAGGA ACATTTCTG TCGTGTACTG TGCGGTGCTT GGGTGGGAC 20510 AGCCAGAGAT GGAGCCACCC CGCAGACGGT CGGGGTGGG CGCTGGCTC GTGTCCTG GAGGGGGAGC 20580 70 TGGGCTGGGC CTGTAAGTCC TCAGCCTCTG TTTCCCTCA GGGATGTGCG TGGGGGCCA GGGCGCCGCC 20650 GGCCCTCTGC CCTCCGGAGG CCTGAGCTGC CGTGTGGCAGG AAGCATTCTC GCTCAAGCTG ACTCGACACC 20720 GTGTCACCTA CGTGCACCTC CGTGGGTGAG TCAGGACAGG CAAGTGTGGG TGGAGGGAG TCGGGGGCC 20790 ACCTGCCAG GGGTCATCT TGAAACCCCT GTGTGGGGCG AGCAGCTCA GATGCTGCTG AAGTGCAGAC 20860 GCCCGGGGGC CTGACCCCTGG CGGGCTGGAG CCACGCTGGG AGGGCTATGT GATTAACACG TGTTGCCCC 20930 AGGCCACCGGA CGCTGGCAGG TCTCCAACT TCTTGAACCC CTGCTTCCCA TCTCAGGGCC GATGGCTCCC 21000 CAGCCTGGGG AGCCTCTGA CCCCTGACCT GTGTGCTCTC ACAGCCTCTT CCCCTGGCTC TGCCCTGAGC 21070 TCTCTGGGGTC CTGAGCAAGT TCTCTCCCG CCCCAGGGCT CCAGCGTCAC TGGGGCTGCT GTCTGCTCC 21140 CCCGGTGGAG GGGTGTCTGT CCCCTACTG AGGTTCCAC CAGCCAGGGC CACGGAGGTC AGGGCCCTGCC 21210
---	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

5	TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTCTGGA ACGGAGTCG 21280 ATTTTGGCC CGCAGCCAG ACGCAGCTGA GTCGGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350 CGCAGCCAA CCGGCACGTC CCTCAGACTT CAAAGACCAC CTGGACTCAT GGCCACCCGC CCACAGCCAG 21420 GCCGAGAGCA GACACCAGCA GCCCTGTAC GCGGGGCTCA ACCTGCCAGG GAGGGAGGGG CGGCCACAC 21490 CCAGGCCGC ACCGCTGGGA GTCTGAGGGC TGAGTGAGTG TTTGCGGAG GCCTGCATGT CGGGCTGAAG 21560 GCTGAGTGTC CGGCTGAGGC CTGAGGGAGT GTCCAGGCA GGGCTGAGTG TCCAGCACAC CTGCCCTT 21630 CACTCCCCA CAGGCTGGC TCCGGCTCCA CCCAGGGCC AGGAGGACG CACAGGAGC CGGGCTTCCA 21700 CTCCCCACAT AGGAATAGTC CATCCCCAGA TTGGCATTG TTACACCCCTC GCCCCCTCC CTTTGCTT 21770 CCACCCCCAC CATCCAGGTG GAGACCTGA GAAGGACCCCT GGGAGCTCTG GGAATTGGG GTGACCAAAG 21840 10 GTGTGGCTG TACACAGGG AGGACCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGAGGT 21910 GCTGTGGAG TAAATACTG AATATATGAG TTTTCAGGT TGAAAAAAAA TCTCATGTTT GAATCTTAAT 21980 GTGCACTGCA TAGACACCCAG TGATGCAAC TACAGAAGCC TGTGAGTGAA CGGGGTGGTG GTCAGTGCGG 22050 GCCCATGGCC TGGCTGTGCA TTACAGGAAT TCTATGAGTG AATGGGGTTG TGTCAGTGCG GGGCCCATGG 22120 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGGAGG AGGGGAAGGA GGGTAGGGG TAGACAGTGG 22190 15 GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCC GAAGGGCAGC AGGGATGCTG GGGGCCACG 22260 TTGGCGGCCG GGGATGATGG AGGGCGCTGC CAGGGATGATGG GGGGTCACG TGCGGTGGCA 22330 GGGGTGTAGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGGAAGCTG GCTGGGCCCC CTCTCCCT 22400 GCCTCCACC TGCAAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTCTAT 22470 GGAGGTTGGG GGCACGGGCA TGACACCATC CTGTTAAAGA TCCAGGATCTC CTCCCTCTGA AGGCCCCAAC 22540 20 TCAGGTTGAA AGTCACATTC CGGCCCTGGC CATTCTCTGA AGAGTAGAC AGGATTCTGA TCTCTGAAGG 22610 GTGGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTCAGGGT GGGCTGGTG GTCCTCTCA 22680 TCTCTCTTATC ATCTCCAGT CTCATCTCTC ATCCTCTTAT CTCATCTCCAGT CTCCTCTTCA 22750 TCTCCAGTC TCATCTGTC TCTCTTACCT CTCATCTCCAGT CTCATCTCTT ATCCTCTTAT CTCCTAGTCT 22820 25 CATCCAGCT TACCTCCCAG GGCGGGTGC AGGCTCGAG TGGAGCTGGA CATACTGCCT CTCTCAGGCA 22890 GAAGGAACTG GAAGGATTTG AGAGAACAGG AGGGGGGGCT CAGAGGAGACG CAGCTTGGG GTGAAGAAAC 22960 AGCCCCCTCT CAGAAGTTGG CTTGGGCCAC AGGAAACCGA GGGCCCTGG TGAGTGGCTC CAGAGCTTCTC 23030 CAGCAGGTCC CTGGTGGGGC TTATGGTAT GGCGGGTCC TACTGAGTC ACCTGGACCA GGGCTTCTGG 23100 TTTGTAGTCA GCCCCGACGT GCCTGGTGTG GGGGGGGGG CTTATGGCCA CTGGATATGG CGTCATTAT 23170 TGTCTGTGCT TCAGAGAATG TCTGAGTGA CAGAGCTTAAT GTGAGTGGT GGGCCAAGTC CACAGACTG 23240 30 GTGCTAATG CACTCTGGTG CCTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT TTTGGCAGCC 23310 GGCCCTGGGG CGCCCTTGGC CTGCAACACTG GAGGGAGGC GGGCGGGGG CGCTGGGGGG ACCACCTCAA 23380 GTGAGGAGTT GGACAGAACAA GGCGGGGAC TTCCCAAGGAG CAGAGGGCCG TGCTCAGGCA CACCTGGTT 23450 TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTCTAC AAAGCTCCAG ATTCCTGTTT 23520 CTCCGGGTGT TTTTTGTTGA AATTTTACTC AGGTTACTT ATATTTTTCTA CAAAGTATT AGACCCCTAA 23590 35 AAAAGGTATT TGCTTGTATA TGCTCTTAAC TCACTAACAC CTACTTTTT TGTCGTGTTT TATTTTATTAT 23660 TATTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTTAGTG CAGTGGCACA GTCATGGCTC 23730 GCTGTAGGCC CAAACCCCCA GCGCTCAAGTG ATCCCTGGGG CTCAGCTTCC CAGACTGCTG GGATTACAGG 23800 TGTGAGGCCC TGCCCTTGGC TGGCACTTTT AAAAAGGACT AGTAAAGGTC AGGCTTCAAGTG GCTTCCACAC 23870 40 CTGTCATCCC AGTAGTTGG GAAGCCGAGG CAGAAGGATT GTCTGAGGCC AGGAGTTGA GACCAGCATG 23940 GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAAG AGTTATCCTGG CGCTGGGGTC CAGCATCTGT 24010 AGTCCCACTG CTGCGGGAGG TGAGTGGGA GAGATGCCCTG AGGCCCCGGAG GTCATGGCTG CAGTGGCTG 24080 TGATGTTACC ATCGCACTCC AGCCTGGGC ACAGAGTGAAC ACCCTGTCTC AAAAAAAA AAAAAAAAG 24150 AAGGAGAAGG AAGAAGAAGA AAGGAAGAAGG AAGAAGAAGG AAGAAGGAG AAGAAGGGAG 24220 45 AAGGAGGCTC GCTAGGTGCT AGGTAGACTC TCAAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290 GGAAAGAAAA ACCCCAGCTC TTGGACTTC CTTAGGCCCTG AACTTCATCT CAAGCAGCTT CTTCCACAG 24360 ACAAGCGTGT ATGGAGCGAG TGAGTCAAAG CGAGAAAGGG AGGAGAACCA GGCAAGGGTG GAGGGCTGTGG 24430 GTGACACCAAG CGAGGACCCCCA TGAAAGGGAG TTGGTTGTTT CCTGCTCTAG CCCACGCTC CTGGGGCTTC 24500 TGCACCTGCT GTCACCGTGC ATGTTGGTGC CAGGTGCCCA CCTGGGAAGG ATGCTGTGCA GGGGGCTTG 24570 50 CAAACCTTGC TGGGGTTCAAG AGGCCCCAGG CACTTTGTTG AGGCACAAATT ACAGCCCCCTC CCCAAAGATG 24640 CCCAACGTCTC TCTCTTGGAA CCTGTAATG TGTCACTGGC AAGGCAGGG CTGGTGAAGG CTGCAGGTGG 24710 AATCACGGCT GCCAGTCAGC CGATCTTAAG GTCATCTCTG ATTATCTGT GGGCTGTATA TGCCACAAAG 24780 GGTGGCTAGA AGTGGAGAGG GGAGGGAGG GAGAGTCAGA GAGGGGAGCT GAGAAGGACCC ACTGGCCACT 24850 GCTGGCTTGG AGATGGAGGA GGGGGTCCCC AGGCAACGAA TGGGGGAGGC CGCTCCATGC TGAAAAGCA 24920 65 AGCAATCTC CCCGGTCTCTG AGGGCACACG GCCCTGCCCA CGCCTCGATT TCAGGCCAGT GGGACCTGTT 24990 TCAGCTGCTTGGCCCTTGGCAAGTGAGTGATGCTGGTTG TGTTCAAGGCA CTAAGCTGCA GTGATTCTGTC 25060 ACAGCAGCAA ATGGAATAGC AGTACAGGGAA ATGAAATACA GGGACAGTTC TCAGAGTGAAC TCTCAGGCCA 25130 CCCCCTGGG 25138
---	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk. (%)	70	60	80	100	100	95	70			80	100	100	60		

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

5' Intron 1: bp 11493-11596 (SEQ ID NO 4);
Intron 2: bp 12951-21566 (SEQ ID NO 5);
Intron 3: bp 21763-23851 (SEQ ID NO 6);
5 Intron 4: bp 24033-24719 (SEQ ID NO 7);
Intron 5: bp 24900-25393 (SEQ ID NO 8);
5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden 10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);
Intron 7: bp 3879-4858 (SEQ ID NO 11);
Intron 8: bp 4945-7429 (SEQ ID NO 12);
Intron 9: bp 7544-9527 (SEQ ID NO 13);
15 Intron 10: bp 9600-11470 (SEQ ID NO 14);
Intron 11: bp 11660-15460 (SEQ ID NO 15);
Intron 12: bp 15588-16467 (SEQ ID NO 16);
Intron 13: bp 16530-19715 (SEQ ID NO 17);
Intron 14: 19841-20621 (SEQ ID NO 18);
20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position 21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGGTCGGCGTCGGCTGGGTTGAGGGCGGCCGGGGACCACCGACATGCGGAGAGCAGCCAGG
CGACTCAGGCCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCGTCGAGGGCCAGGCCAGAGCTGAATGCAGTAGGGCTCAGAAAAGGGGGCAGGCAGGCC
CTGGTCTCTCTGCTCCATCGTCACGTGGCACACGTGGCTTCGCTCAGGACGTCAGTGACACGGTATCTTGCC

TCTGCTCTCCCTCTGTCCAGTTGCATAAACTTACGAGGTTCACCTTCACGTTGATGGACACGCCAGGGTCCAGGCC

10

CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGCGCCAGTGGAGCCGGTTGCCGCAATGGGAGAAGTGTCTGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTTCGCAATTCAAGGGTGGAAATGAGAGGAGG
CGAGAACCCCTCTCCTGGGGTGGGAGGTAAGGGTTTGAGGTGCACGTGGTCAGCCAATATGCAGGTTGTGTTA
AGATTTAATTGTGTGTTGAGGCCAGGTGGCTCACGCCGTAATCCAGCACTTGGGAAGCTGAGGCAGGTGGA

15

TCACCTGAGGTCAAGGAGTTGAGACCAGCCTGACCAACATGGTAAACCCATCTGTACTAAAAATACAAAATTAGCTG
GGCATGGTGGTGTGCTGTAATCCAGCTACTTGGAGGCTGAGGCAGGAAATCACTGAACCCAGGAGCGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGCGACAAGAGTGAACACTCTGTCTTAAAAAAAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCATTTAGGT
GAAGAGGGCACATGGGAGCAGAGCACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGTGGGTGTTCA

20

ATGGTGTGCTGGCCCTGCCGTGTCCCCACCCCTGTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCC
TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCGATTTCACCCCTCCCCACAAACTCCAAAGAC
ATGTAAGACTCCGCCATGCAGACAAGGAGGTGACCTTCTGGGCTCTTTCTTTCTTTCTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCTAACACCGTTTCTGTGTACAGTGCAGAATTGCTAACTCGCGGTGTTA
CAGCAGGTTGCTGAAATGTCGCTTGCCTGACTGGAAGTCCCTACCCATGAAACGGCAGCTGCCACACTGCTGC
GGCTCAGGTGGACCACGCCAGTCAGATAAGCGTCACTGCAACCCAGTTTGTCTTGTCTCCAGCTCCCTCGTGTGAG

25

GAGAGTTGAGTTCTCTGATCAGGACTCTGCCGTGATTGCTCTGACTTCAGATGAGGTACAATGCCCC
CTTATGCAAGGGACTGAGGCGGTGCCCCGGGTGCTCCCTGTCACGTGCAAGGTGAGTGAGGCCTGCCCC
GTCACTGTAGGGTAGGTGAGTGAGGCGGCCATCCCGGGTGTCCCTGTCACGTGTAAGGTGAGTGAGGCCTG
CCCTGTCACGTGAGGGTAGGTGAGTGAGGCGCTGCCCCGGGTGTCCCTGTCACGTGCAAGGTGAGTGAGGCCTG
GTGTCCCTGTCCCGTGCAGGGTAGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCACGTGCAAGGTGAGTGAGGCCTG

30

CCGGGTGCTCCCTCAGGTGAGGGTAGTGAGGCGGCCAGGGTGTCCCTGTCACGTGTAAGGTGAGTGAGGCACC
GTCCCTGGGTGTCCTCCAGGTATAGGGTAGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCACGTGCAAGGTGAGTGAG
CGCGGCCCCCGGGTGTCCCTCAGGTGAGGGTAGTGAGGCGCTGCCCCGGGTGTCCCTGTCACGTGTAAGGTGAGTGAG
GAGGCTGTCCCCAGGTGTCTGGCTTGCTCACTTGAGCTGCTCCCTGTCACGTGTAAGGTGAGTGAGGCACC
GCGCCGGTTGCCATTGCTGGTAGATGGTAGCAGGCCAGTGCTGGCCCCAGCCTATCTTGATGCTGCCCT

35

TCTTGGTCACCTCTCCGTTCCATTGCTACGGGACACGGACTGCAGGCTCTGCCCTCCGCGTGCAGGCACAGC
CCACAGCTTCAGGTCCGCTTGCCTCTGTGGGCTGGCTGTCACCGTGCCTGCCACATGCATGCTGCCAATACTCC
TCTCCCAGCTTGCTCATGCCAGGCTGACTCTGGCTGCTGTGCTGCTGCCACGTGTTGCTGGAGACATCCAGAA
AGGGTTCTGTGCCCTGAAGGAAAGCAAGTCACCCAGCCCCCTCACTGTCTGTTCTCCAAAGCTGCCCTCTGC
TTGGCCCCCTGGGTGGCAACGCTGTCACTTATTCTGGCACCTGCCGCTATTGCTTAGGCTGGCTCTGCC
40 CCAGTCGCCCTCACATGGATTGACGTCCAGGCCAGGGTAGTGAGTGCTCTGTCTGCTCTGCCAGGACAGTG

GAGGGCCGGTCTCCGCCAGCCTCGTCAAGACTCCCTTGGGTCTTAGTTGAATTCACTGATTACCTCTGACG
 TTTCTATCTCCATTGTATGCTTTCTGGTTATTCTTCATTCTTCTAGCTTCTTAGTTAGTCATGCCCTC
 CCTCTAAGTGCCTAACCTGCAACCTGTGTTGATGTGAAGTAATCTCAACATCAGCCACTTCAGTGTCTTAA
 ATACCTCAAGTGTAAATCTTTAAGTATTCTTCTGTGATTTTCTTGTGCAGCCTGTGTTGACGTGA
 5 AATCATTTGATATCAGTGACTTTAAGTATTCTTAGCTTCTGTGATTCTTGTGAGCTGAGTTATTGAACACT
 GTTATGTTCAAGATATGTAGATCAAGATACGTAGAGTATTTAAGTATTCTAATTGATTCTAACTCAGT
 TGTGAGTGGTCTGTATAATACCAATTATTGAAGTTGCGGAGCCTGCTTGTGATCTAGTGTGCATGGTTCCAG
 AACTGTCCATTGTAATTGACATCCTGCAATAGTGGCATGATGTCACTATATCCAGTTATTAGGTCCAGTGA
 AAGCTCTGTCCTCTAGATGCATGAAATTCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGATGGTCTGTTATT
 10 TCTTCGTTGGTAGCTTATGTCTAGTAATTCTAGTAATTCTTTAAATTGCTCTAGTACTGCACACTGGCTCT
 TCTTTGGAGACTTCTATGTCTAGTAATTCTAGTAATTCTTTAAATTGCTCTAGTACTGCACACTGGCTCT
 TTTGATTAGTATTTCTGCTGTCTGTTCTGCCTTAATTATATATATATATATTTTTTTGAGACA
 GAGTCTTGGCTGTCGCCAGGGTAGTGCAGTGGTGTGATCACAGGTAGTGAACCTTACCTCTGGCTGAGCCGT
 CCTCTCACCTCAGCCTCTGAGTAGCTGAACTGCGACACCGCACCGTACACCTGGTAATTAAATTCTGGTA
 15 GACAGGGTCTGCTGTGTTGCCAGGCTGGTCTCAAACCTTGGACTCAAGGGATCCATCACCTGGCTCCAAAGTG
 CTGAATTACAGGCATGAGCCACCATGTCGGCTAATTCAACACTTTATATTCTTATAGTGTGGTATGCTGTTA
 ACAGCATGTAGGTGAATTCCAATCCAGTCTGACAGTCGGTAACTGGATAACCTGATTATTCACTTTGGTCA
 ACTAGAGACCCGCTGGTCACTCTGATTCTCACTTGGCTGGTCACTGCTCTCGTCCCTGGTCACTGGCATT
 20 GGTGCCATGCGTTCTGCCAGTGTGTTGATCCTCTGGCTCTGGTCACTGGCTTCTGGTCACTGGCTTCT
 CTTGCTTAGTGTACCCCCGATCTTTATTGCTGTTGCTTTGTTATTGAGACAGTCACTCTGTCACCA
 GGCTGGAGTGAATGGCACAATCTGGCTCACTGCAACCTCTGCTCTGGTCAAGCAGTCTCATCCCAACCTCA
 TGAGTAGCTGGATTACAGGGCCACACCACGGCTGGTAATTGGTATTAGTAGAGATAGGCTTCAACCATGT
 TGGCAGGCTGGCTCAAACCTCTGACCTCAAGTGAATCTGCCCCCTGGCTCCACAGTGTGGGATTACAGGTGAA
 25 GCCACCGTGCCGGCATACCTGATCTTTAAATGAAGTCTGAACATTGCTACCCCTGCTGAGCAATAAGACCC
 AGTGTATTTAGCTGGCACCCCCCAGCCTGTGTTGCTTTCCCTGACTTAGTTCTATCTCAGGCATCTGACA
 CCCCCACAAGCTAAGCATTATAATTGTTTCCGTGAGTGTGTTCTGAGCTTGCCCCGCCCTGCTTTCTCC
 TTGTTCCCCCTGCTCTCTGCTCAGGCCGCGCTGGGCTCCCTGCTGGCTTCTGCTGGTGGTCTCTGCTT
 TTATTGCTGGTAAACCCCGCTTACCTGCTGGCTCCATGGCATCTAGCGACGTCGGGACCTCTGCTTATGTC
 ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGGTCACTTGGCCCTGGTCACTGGAGCACCACTGGCCAGCGT
 30 CTTAGCCAGTGAATGACAGCAACGTCGCTGCCCTGGTTCAAGCTGGAAACCCAGGCATGTCGGGTCTGGGCT
 CGCGGTGTGAGTTGAATCGCAGAACCTGCGGTGTTGGCGCAGCTCTGACGGTGTGCTGCCCTGGGGAGTGTCTG
 CTTCTCCCTCTGCTGGACCAGGACAAGGATGAGGCTCGAGGCCAACAGGAGCATGACGTGAGCC
 ATGTTGATAATTAAAATTCTAGGCTGGCGCGTGGCTCACGCCCTGTAATCCAGCACTTGGAGGCCAAGCGGG
 TGGATCACGAGGTAGGAGGTCAGAGGAGCATCTGGCAACATGATGAAACCCATCTGACTAAAAACACAAAATTAGC
 35 TGGGCGTGGTGGCGGGTGCCTGTAATCCAGCTACTGGAGGCTGAGGAGGAATTGCTGAACCTGGAGTTGAA
 GTTGCAGTGAAGCGACATTGACCACTGCACTCCAGCCTGGCAACACAGCAGACTCTGCTCAAAAAAAA
 AAAAAAAAATTCTAGTAGGCCACATTAAAAAGTAAAAAGAAGGGTAATTATGTAATAATAGATTTACTGAA
 GCCCAGCATGTCACACCTCATCTTGGTGGAGCATCACTCACAGGACATTGACATTGACATTGAGCT
 TTTGCTGGGATCCCGTGTGAGGCTGGCGTGGCATCTGCCCTGGACCTGCTGGCTCCATGCCATGGCT
 40 GTTGTACCAAGATGGTGCAGGTCGGGAGGCTGAGCTGGATGTGAGCTGGATGTGAGCTGGAGGGTCA
 TCTGGGATGAGGTGCCAGGCCCTGCTGAGCTGGATGTGAGCTGGATGGTCTGGATGGTCAAGGGTGAGGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGAT
 TGTGGTGTCTGGATGGTCAGGTCAAGGGTGAGGTCTCAGGCCCTCGTAAGCTGGAGGTATGGAGTCGGATGATGCA
 GGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAG
 AGGCCCTCGGTGAGCTGGTGTGCCAGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAGACGGTGCCAGACCATGC
 5 GGTGAGCTGGATATGCCGTGTCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCT
 GTCCGGATGCTGCAAGTCGGTGTGAGGTCAACGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTG
 GGGGTGAAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAG
 GCCCTCGGTGAGCTGGATGTGCACTGTCAGATGGTCAGGTCCGGGTGAGGTGCCAGACCCCTGCCAGGCTGGATGG
 10 TGCGGTGTCTGGATGGTCAGGTCTGGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTACAGGTCTGGAGTG
 GGTCCGGGTGAGGTGCCAGACCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTACAGGTCTGGAGTGAGGTGCC
 AGACCCCTGCTGTGAGCTGGATATGCCGTGTCGGATGGTCAGGTCAAGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGG
 GGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAACTGGATGTGCCGCTCTGGATGG
 15 GCAGGTCTGGGTGTGGTCGCCAGGCCCTCGGTGAGCTGGAGGTATGGAGTCGGATGATGCCAGGTCCGGGTGAGGT
 CCAGGCCCTGCTGTGAGCTGGATGTGCCGCTCTGGATGGTCAGGTCTGGGTGTGGTCAGGTCCGGGTGAGGT
 GAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCCGCTCTGGATGG
 GTGCAGTCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTACAGGTCTGGAGTG
 20 CACCAAGGCCCTGCGGTGAGCTGGATGTGCCGCTCTGGATGGTCAGGTCTGGGTGTGGTCAGGTCCGGGTGAGGT
 TGGATGTGCCGTGTCGGGTGAGGTGCCAGGCCCTGGTGTGGTCAGGTCCGGGTGAGGTCTGGATGTGCCGCT
 GTCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGGTGTGGTCAGGTCTGGGTGAGGT
 GGGTGAGGTGCCAGGCCCTGGTGGCTGGATGTGTGGTCAGGTCCGGGTGAGGTGCCAGGCCCT
 25 GCTGTGAGCTGGATGTGCCGTGTCGGATGGTCAGGTCCGGGTGAGGTAGCCAAGGCCCTCGGTGAGCTGGATGG
 GTGCCGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTCGGTGAGGTAGCTGGATATGCCGGATGGTCAGGT
 CCAGGCCCTGCGGTGAGCTGGATGTCTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCA
 30 CGCTGTGAGCTGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGGTGTGGATGGTCAGGTCTGGGTGAGGT
 TGGATATGCCGTGTCGGGTGAGGTGCCAGGCCCTGGTGTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGG
 GTCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGGTGTGGATGGTCAGGTCCGGGTGAGGT
 GGGTGAGGTACCAAGGCCCTCGGTGATCTGGATGTGCCATGTCTTCGTTAAG
 35 **Intron 3 (SEQ ID NO 6)**
 GTACTGTATCCCCACGCCAGGCCCTCTGCTTCTGAAGTCCTGGAAACCCAGGCCGCCCTCAGCATGCCCTGTCTCCACT
 TGCCCTGTGCTTCCCTGGCTGTGCAAGCTCTGGCTGGAGCCAGGGGCCCGTCACAGGCCCTGGTCCAAGTGGATTCTGTG
 CAAGGGCTCTGACTGCCCTGGAGCTCACGTCTCTTACTTGAAAATCAGGAGTTGTGCCAAGTGGTCTTAGGGTTTGTA
 AAGCAGAAGGGATTAAATTAGATGGAAACACTACCAACTAGCCTCCTGCCCTGGGATGTGGGTCTGATTTCTC
 40 TCTCTTTTTTTCTTTTGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCTGGTTAACCGATTACCCAGCCTCAGCCTCTAACGTAGCTGGATTACAGGCACCTGCCACAC
 GCCTGGCTAATTTTGTACTTTAGGAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACCTCATGACCTCAGG
 TGATCCACCCACCTGGCTCCAAAGTGTGGGTTACAGGCTAACGCCACCGTGCCAGCCCCGATTCTCTTTAATT
 CATGCTGTTCTGTATGAATCTTCAATCTATTGATTTAGGTATGAGAGGATAAAATCCACCCACTGGGACTCACTG
 5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGATAGGAGAGTTCCACCATGAGCTAACCTCTAGGTGGCTGCATTGAAATG
 GCTGTGAGATTGTCTGCAATGTTGGCTGTGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGATCAGTGAG
 GGACGGGAGCCGTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTCTCCGTGCCCCGCCAGG
 TGACTGTGGAGGGCTTAGTCAGAAGATCAGGGCTTCCCAGCTCCCCTGACACTCGACTCCCTGGGGGCTTGTGAC
 10 ACCCCATGCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
 GACCCGACGTGGTCTGGGCCATTCTGCATCTGGGGAGGGTCAGGGCTTCCCTGTGGAAACAAGTTAACAC
 ATGACACCTTACTTAGACTTACACGTATTAATGGTGTGCGACCCAACTGGTCATTGACCAGTATTTGAAAGAAT
 TTAATTGGGGTGAACCGAAGGAGCAGACAGACGTGGTCCCCAAGATGCTCTTGCACTACTGGGACTGTTGTTCTG
 CCTGGGGGCTTGGAGGCCCTCCCTGGACAGGGTACCGTGCCTTCTACTCTGCTGGGCTGCGGCTGCGGCT
 15 AGGGCACCACTCGGAGCACCCGGCCAGTGTCCACGGAGTGCAGGCTGTCAAGCCACAGATGCCAGGTCCAGG
 GTGGCCGCTCAGCCCCCGTCCCCCATGGTGGTTGGGGAAAAGGCCAAGGGCAGGGTGTCAAGGAGACTGGTGGG
 CTCATGAGAGCTGATTCTGCTCTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCCCTCCATCTGAAGGGATGTGGCT
 CTTTCTACCTGGGGTCTCTGGCTGGGGCAGCCTGGGCTACCCAGTGGCTGTACCAAGGGACAGGCATCTGTG
 AGGGGATGGGTTACGTGGCCCTGGGACTATAGGACCAGGTGTCAGGTGCCCCGCAAGTAGAGGGGCTCAG
 20 AGGCGTCTGGCTGGCATGGTGGACGTGGCCGGCATGGCCTTCAGCGTGTGCTGGGTGGCTGAGCCCTCAC
 TGAGTCGGTGGGGCTTGTGGCTCCGTGAGCTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCTGAGGGGCTCT
 CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTGGTTAACCTCTTTAACAGAACAGTGTGTTGAGCCCCACATTGGTATCAGCTTAGATGAAGGG
 CCCGGAGGAGGGCACGGACACAGCAGGGCATGGCACGGCGCAACCCATTGTGCGCACAGTGAGGTGGCCAGG
 TGCCGGTGCTCCAGAAAACAGCGTGGGGTGTAGGGGGAGCTCTGGGGCAGGGCAGGGCTCTGAGGACCACAAGAAG
 CAGCCGGGCCAGGGCTGGATGCAGCACGGCCAGGTCTGGATCGTCTGCTGTGGTGCAGCCTCCGTGCGCT
 TCCGCTTACGGGCCCCGGGACCAGGCCACGACTGCCAGGAGCCCACGGGCTCTGAGGATCTGGACCTGGCCACGG
 30 CTCCCTGCACCCACCCCTGTGGCTGCGTGGCTGCGGTCAGGGCTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG
 GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGAACCCGTTCAACACAGGGCTGAGGAAGCTGGGGGG
 TTCTAGGTCCCGGGCTGGTGGCTGGGACACTGGGAGGGCTGCTCTCCCTGGTCCCTATGGTGGGTGGCAG
 TTGGCCGGATCCACTTCTGACTGTCTCCATGCTGTCCCCGCCAG

Intron 5 (SEQ ID NO 8)

35 GTGGGTGCCGGGACCCCGTGAGCAGCCCTGCTGGACCTGGGAGTGGCTGCCCTGATTGGCACCTCATGTTGGGGAG
 GAGGTACTCTGGGTGGGCCAGGGAGTGCAGGTGACCCCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
 CTTCAGCCTTCCCTGCAGCACATGGGCCAGTGTGACCCCTGACTGCCCGGGCTCTTATCCAAAGGGGGTCCACTG
 GATTCCAGTTCCGTAGAGAACCGAACGGCTCAGCCACCAGGGGGCTGCTCTCCCTGGTCCCTATGGTGGGTGGCAG
 40 GGGTCTCCTGCTCTGAGGCTCAGAGAGGGACACAGCCGCCCTGCCCTGGGTCTGGAGTGGTGGGGTCAAGAGAG

AGTGGGGGACCCGCCAGGGCAGGGCCCTGAGGGCAGAGGTGATGTCTGAGTTCTGCGTGGCCACTGTCAGTCTCCCTCGC
CTCCACTCACACAG

5'-Bereich Intron 6 (SEO ID NO 9)

5 GTAAAGGTTACCGTGTATAGTCGTGTCAGGATGTGTCTCTGGATATGAATGTGTCTAGAATGCAGTCGTGTCGTG
ATGCCGTTCTGGGTGGAGGTACTTCCATGATTACACATCTGTGATATGCGTGTGGCACGTGTGTGTCGTGTCAT
GTATCTGTGGGTGCATATTGTGGTGTGTGTGTGGCACGTGTGTGTCATGGTGTGTGTCCTGTGGTGTGCAT
TGTGTGTGTCTGTGACACGTGCATGTTCATGCTGTGTGCGATGTCTGTGATGTGCTATTGTGGTGTGTGTCAT
GTGTCGTGACATATGCGTGTCTATGGCATGGGTGTGTGTCGCCCTGGCCTACTCCTCCTCCAGGCATGGTCC
10 GCACCAATTGTCTCACGCTCTGGGTGCTGGTTGGGAGCTCCACATTCAAGGGCTCTCACTCTAGCATGGGTGCCCT
GTCTGTCAACAGGGCTGGGCCCTGGAGACTGTAAAGCCAGGTTGAGAGGGAGAGTAGGGATGCTGGTGTACCTTCCTGG
CCCCCTGGCACCCCCCAGGACCCCAGTCTGGCCTATGCCGCTCCATGAGATATAGGAAGGGTGTGATTCAAGGCCCGCT
GGGACACACTCCTCCAGAGCGGCCGGGCCCTGGGCTCGCAGGGTGAAGGGGCCCTGGCTGGTTCCACCCACCC
15 AGTGGTCATGAGCACGCTGGGGGTAAGCCCTAAAGTCGTGCCAGGGCGGGGTGAGAGGTGAAGAAGTATCCCTGG
GCTTCGGTCTGGGGAGAGGCACATGTTGAAACCCACAAGGACCTCTTCTGTACTCTTGAGGCT

3'-Bereich Intron 6 (SEQ ID NO 10)

Intron 7 (SEQ ID NO 11)

30	GTCTGGGCACTGCCCTGCAGGGTGGGCAAGGACTCCCAGCAGTGGGCTCCTCCCCTGGGCAATCACTGGGCTCATGACCC GACAGACTGTTGGCCCTGGGGGCACTGGGGGAATGAGCTGTGATGGGGCATGATGAGCTGTGCCTGGCAAAATC TGAGCTGGGCACTGCCAGGCTGCCACAGCTGCTGCATTAGGCACCTGCTCACGTTGACTGCACGCCCTCTCCAGTT CCGCACTGCCCTTGTTCATGTTGCTAAATGCTTCTCTGCCAGTTGATCTTGAGGCCAAAGGAAAGGTGTCCCCCT CCTTTAGGAGGGCAGGCCATGTTGAGCCGTGCTCTGCCAGCTGGCCCTCAGTGTGGGCTGAGGCCAAAGGAAACG TGTCCCCCTCTTAGGAGGACGGCCGTGTTGAGCCACGCCCGCTGAGGGCCTCTCAGTGTGGGCTGTCCACGT GGCCCTGTGCCCTTGAGATGTGGTCTGCCACGTGGCCCTGTGGCTTTGAGATGCTGTAGCACTTGCTCGG 35
35	TCTAGGGACAGTCGTGTCACCGCATGAGGCTCAGAGACCTCTGGGCAATTCTGGCTCCAGGGTGGGGTGGAG GTGGCCTGGCTGCTGGGACCCAGACCCCTGTGCCCGCAGCTGGCAGCAACTCTGGATCACATATGCCATCCGGCCA CGGTGGCTGTGGGTGAGCCAGCTGACCCACAGTGGCCAGAGGAGACCTCTGTGTCACACACTCTGCCAA GCCCATGTGTCTGCAGAGACTCGGCCCGCCAGCCCACGATGGCCCTGCATTCCAGGCCAGCCCCGCACTTCATCACA AACACTGACCCAAAAGGGACGGAGGGCTTGGCCACGTGGCTCTGCTCAGCACCCACGGCTCACTCCCATGTG 40
40	TCTCCCGTGTGCTTCGCGAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTCGGGTGGCTGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCC
 CTTCCCCACTGNCTTCTGCCCGGGCCACCAGAGTCTCCTTTCTGCCCTCCGGCTCTGGCTCAGGTGCGCCAGCGCTGGAAATGGCAAGC
 5 TCCCGAGGCCCGAACATGGCTCGGCTTGCGCGAGCCGGAGCGAGGGGGGGGGGTGTGCTGGTCAAGGTGCGCCAGCGTTGAGCCT
 GGGGTGAGGGATTGCTCTCGCTGGAGGACGAGGGGGGGGGGTGTGCTGGTCAAGGTGCGCCAGCGTTGAGCCT
 GCAGCTTGTCAAGCTCAAGTTACTACTGACGCTGGACACCCGGCTCACACGCTTGTATCTCTCTCCGATAACAAAA
 CGATTTATCCGATTCTCATTCCTGTCCCTGTCGTGACCCCCCGAGGGCGCGGGCTTCTCTGTGACTAGATT
 CCCATCTGAAAGTGGGGGGTGTGACCGTGTAGTTGCTCTCGGGGGCTGTGGTGGCCATGGGCAGGCGGCC
 10 GAGAGCTGCCGTACACAGCCACTGGGTAGGCCACACTCACGGTGGTAGAGGCCACAGTGCCTGGTGCACATCACGTCT
 CTGGATTTAAGTAAACACACACCTCCCGCAGGCATCTGCCCTGCACCCCTGTGTGCGCTGGGAGAGTGGTAGCAC
 CGAGGAAATTCGTGCACACTCAAGGTCAAGGTCAAGGCACAGGGCTCTCGCAGTCAAGGTGGAACTGGAGGCTCTCTGGGATC
 GTCTCCAGCGATAAAGGACTGTGACAGCTCGGAAGCTTTATTTAAATAACTATTAAATTATTGCAATTATAAGT
 AATCACTAATGGTATCAGCAATTATAATTATTAAAGTATAATTAGAAATATTAAAGTAGTACACACGTTCTGGAAAAAA
 15 CACAAATTGCACATGGCAGCAGAGTGAATTGGCCGAGGGACACGTGTGCACATGTGTGAAGCGGCCAGGCCAC
 AGAATTGCTGACAAAGTCACCTCCCCAGAGAACGCCACACGGGCCCTCTCGTGGTGTGAATTATTAAAGATGGATC
 AAGTCACGTACCGTCCACGTGTGGCAGGGCTTGGGGAAATGTGAGGTGATGACTGCCTCATGCCCTGACAGACAGGA
 GGTGACTGTGCTGTCCCTGCTCTAGGACACGGACAGGGCCGAAGCTAGTCCCCTAGTGGTCCAGTTGGCCTCTGA
 ATAAAAAACGTCTTCAAAACCTGTTGCCCAAAAACAAGAACAGAGAGAGTTCCATCCCTGTGTCACAGGGCGTA
 20 TCTGCTTGCCTGACTCGCTGGCTGGCCGACTCTAGAGTTGGTGGCTGTGCTCTGTGCAAAAGTGCAGTCCTCTT
 GCCCATCACTGTGATATCTGACCAAGCAAGGAAAGCCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 CTGTTGCTGCTGGCTTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGGCTCCGGGTTCCAGCATTCTC
 CTGCCCTCACCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCCCTGGTAATTGGTATTTAGTAGAGAG
 GGGTTTTGCCATGTTGCCAGGCTGGCTCGAACCTCGACCTCAGGTGATCCACCCACCTGGCCTCCAAAGTGTG
 25 GGATTACAGGTGTGAGGCATCACGCCAGCCGGAAAGCCTTTTAAGGTGACCACCTATAGCGCTTCCCAGGAAATAAC
 AGGTCTTGTGTTTGAGTCAAGGCTCTTAGCAACAGGAGTGGCGTCTGTGGCTCTGGGATGGCTGAGG
 TCGCGTGGCAGCCATGCCCTCTGTGCACTTTAGGTTCCACGGGCTATTCTGCTCACTGTTGTGAAACAGC
 CCCTTGGCATCTTGTGAGGAGTTCTGCTGTTGAGTCAAGGAGTGGCGTCTGTGGGATGGCTGAGG
 GCAGCGGCTACATGTAGGGTCACTGAGTTACCGTGACAAATTCTGAAAAAGGAGTCCGGTTAACAT
 30 TCATTCCGGTCAAGTGTCTGGTCTGTGAATAAAACTCTAACAGTTAACAGGAAACCTTAATGAAAGAAAACCTTGTGATT
 AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTCAAGCCGGCCAGTGCATGGTGGAGAGTGGGGAGCAGGGATTG
 TTTGTTCAAGGGTCTCATCTGGTATGTTCTGAGGTGTTGCGGGCTGAATGGTAGACGTGTCGTTGTGTTGAGG
 TCTGTTGCTGTGTTGAGGCTGGCTGGCTGGTTGAGTGTACGCATGCCAGCACATGCCCTGCCCTCTCACCTGTCTTCCCG
 CCCAG
 35

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCTCTTCCCAGGGGGCTTGGTGGGGTTGATTGCTTTGATGCATTCAAGTGTAAATTCTGGTGC
 TCTGGAGACCATGACTGCTCTGCTTGTGAGGAACCAGACAAGGGTGCAGCCCTCTGGTATGAAGCCGACGGGGAGGG
 40 TTGCACAGCTGAGGACTGCGGGCTCACCGAGGCTCTGCTCCAGGGCATGTCCAGGGCTCAGGGCTCAGCAGGG
 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTCTGTACGTACGTACCCAGGTT

CGTTAGGGCTTGGGGAGATGGGCTGGTGAGCCTGAGGCCACATCTCCAGCAGGCCCTGACAGGTGGCTGGA
 CTGGGCCCTTCACTGCCCCATCCACTGCATGGGCTACACCCAAGGACGACACACCTAAATATCGTG
 AACCTAATGTGGTCAACTCAGCTGTTTATTGACAGCAGTTACTTTTTTTAACTTTAAGTTCTAGGGTAC
 ATGTGACGACGTGCAAGTTAGTTACATGTATACTATGCACTGCAATGGGTGTGCTGACCCATTAACTCATCATTACA
 5 TAGGTATATCTCTAATGCTATCCCTCCCCTCCACTCCCATGACAGGCCCTGGTGTGATGTTCCCCACCCCTG
 TGTCCAAGTGTCTCATGGTCAGTCCCACCTGTGAGTGAGAACATGTGGTGTGGTTCTTCCTTGCAATAGTT
 GCTCAGAGTGTGGTTCCAGCTCGCCATGTCCTACAAAGGACATGAACCTCATCCTTTTATGACTGCATAGTATT
 CCGTGGTGTATATGTGCCACATTTCTTATCCAGTCTATCATGATGGACATTGGGTGGTGCAGTCTTGCTACT
 GTGAATAGTGGCGCAATAAACACATGTCATGTCATGGTCTTATAGCAGCATGATTATAACCTTTGGGTATACCCAG
 10 TAATGGGATGGCTGGTCAATGGTATTTCTAGTCTAGATCCTGGAGGAATCACCACACTGTCCTCCACAATGGTGAA
 CTAGTTTACACTCCCACCAACAGTGTAAAGTGTCTGGTGGAGGGATGTGGACAGCAGTTATTTTTATGAAAAA
 TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTCTCGAACAG
 TCCGGGTTTCTGTGCATCTTGAACTCTAGCTCAATTATAGCATGTCAGTGGATCAAGGTTCTTCATTAA
 15 GGTTCAAGTCTAGATTGAAATAAGTTATGTAACAGAAACAAAATTCTTGACACACAATTGCTCTGGGATTGG
 GGAAAGTCTCGAGCTGGCGCACACTGGTCAGCCCTGGGACAGGATACCTCTGGCCATGGTCATGGGCGCTGG
 GCTTGGGCTGAGGGTCACACAGTCACCATGCCAGCTCTGTGGATAGGATCTGGGCTCGGATCATGCTGAGGACC
 ACAGCTGCCATGCTGTAAGGGCACCACGTGGTCAGAGGGGGCGAGGTTCCAGGCCAGCTTCTACCGTCTTCAG
 TTATTTCTCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCTCGTCTCGACTGGCACAGAATTGACACAA
 GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGATTGCTGAGCTTAAGTGTAGAGAGCTGCTGTGGAAA
 20 GAAATTAAAGTTTCTTAAACCGCTTGGAGAATGTTACTTTATTTATGGCTGTGAAATTGTTGACATTCACTGCC
 TCGTAGACAGATACTACGTAAAAGTGTAAAGTTAACCTTGTGTTGTTGACATTCCCTTATTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCGCGTGTCTGTGGGACCTCCACAGCCTGTTGAGGCCCGTGTCTGCCCTGG
 25 CACCGCAGCTTGTCTGCCAAGTCCTCTCTGCCGGTGTGGATCCGCAAGAGCAGAGGCCCTGGCGTGCACC
 CAGGCCCTGGGGCGCAGGGCACCTCGGGAGGTGGTACCGTGAGGCCCTGGTGCAGAGACGCCACCGG
 ACACACGTGGTGAGTGCAGCGGTGACCTGGCTCTGCTGCTTGGAAAGTCAGAGTGGCGCTCTGGGGCCAG
 TGAGACCCCCAGGAGCTGTGACAGGCCCTGAGGGCCAGGGGGAGCCTCCCTCCAGGGTGACCTGAGCCTGG
 GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTCTGCGGTACGTTCTGCGTGGGTGTTGGGATCGGTGG
 30 AGAATTGGATTGCTGAGTGCTGTCTGAACCACGGAGATGGCTAGGAGTGGTTCAAGAGTTGAAATTGTAAT
 CAAACTAAAATCAGGCACAGGGGACCTGGCTCAGCACAGGGATTGCAATGTGGCCCCCTCAAGGGCCACAG
 AGCCGGTGGCTTAAAGTGCATTGACGAGGGACGAGAAACCTGAAAGGGAAACCTCAGAAAATG
 TGGCCGCCAGGGTGGCTTCAAGGTCTTGTGGCTGTGGTAAACCCATTGGACCCCTCCAAGTCACCC
 TCCAGGTCCACCCCTCCAGGCCGCCCTGGCTGGGGTATGCCCTGTGCGCAGCCGGAGCACAGCAGGC
 35 TGTGACATTTAAATCCACTAAGATTCACTGGGGAGGCCAGGTCCAAGCAACTGAGGCTCAGGAGTCTGAGGCT
 GCTGAGGGACAGGCAGACGGGAACCTGCTCTGTCAGTTCTGAGGCTGCTGGCCAGGGAGCTGGCTCAGA
 GTGTATGTTGGGCTCCACGGGGCAGAACACTCTGTCTGTGAGTCGGCAGCCATGAAACAGGAAGGGTGGCCACAG
 GGAGCTGGGAATGCACCAAGGGAGCTGGCAGCTGGCCAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCC
 GGGGCCACAGCAGGGCCAGGAAGGGAGTCCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG
 40 AGCTGGGTGAGCGAGGCTCATGACTCGGGAGGGAACCTTGTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACCTCAGAACCCCTCCCTTGCTAAAGCACAGCAGATGCCCTCAGGGCATCT
 AGGAGAAAACAGGAAAGTCCTTGAGAACGCTCTAAAGAAGGTGGATGGTGGCAATTCTGTCCAGATTAGTCT
 GCCCCGGACCACAGATGAGTCTATAACGGGATTGTTGCTGACATGGGGACACATGAGATGGACCATCACAGAGGCCAC
 TGGGGCTGCACCTCCCATCTGAGTCTCTGGCTGTCGGGGTCAGGGCAGGTTCTGCATGCTCACCTACCTGTCTGCC
 5 GGGAGACAGGAAAGCACCCGAAGTCGGAGCAGGGCTGGGTCAGGCTCTCAGAGCTCTGCCAGGGCCAGCACCC
 GCTCCAAATCACCACTTCTGGGTTTCCAAAGCATTAAACAAGGGTGTCAAGGTTACCTCTGGGTACGGCCCGCA
 TCCTGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCGGAAGTGGAGCCTGTGCCCGCTGGGGCAGGTGCTGCTGCAGGGCCTTGGTCCACCTCTGCT
 TCCGTGTGGGCAGCGACTCCAATCCAAAGGTCAAGAGGCCACAGGGTGCCATCTGGGCTGAGCAGAATTTGTCTATTTGGTAAAGGAAATGGTGCAC
 AATGCATTTCTGTGGAGGTGAGGGTCTCACACGGGAGCAGTTCTGTCTATTTGGTAAAGGAAATGGTGCAC
 CAGACCTGGGTGCACTGAGGGTCTTCAGAAAGCAGTCTGGATCCGAACCAAGACGCCGGGCGCTGCTGGGCTGAGT
 CTCTCAAACCGAACACAGGGCCCTGCTGGCATGAGTCCCTCTGAACCCGAGACCTGGGCGCTGCTGGGCTGAGT
 15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
 AGTCTACAGGATGCCATGAGTTCATGATCACGTGACCCATCAGGGACAGGGCATGGTGTGGGGGGTCTCTACAA
 AATTCTGGGTCTTGTCTCCAGAGGCCAGAGCTCAAGGCCCGTCTCAGGCTCAAGACACAAATGAATTGAAGATGGA
 CACAGATGCAGAAATCTGTGCTGTTCTTATGAATAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
 ATAATCCCAGCACTTGGGAGGCCAGGGTGGGATCACTTGAGGCCAGGAGTTGAGGCCACCTAACCAACATAGT
 20 AAATTCCATTCTACTTAAAAAAACAAAAATTAGCCTGGCTGGCAGACGCCCTGTAGTCCCCCTATGCCGGAGGC
 TGAGGCAGGAGAACATTTGAACCCAGGAGGAGGGTGTAGGCTGAGGCCAGAGTACACCCACTGCACTCCAGGCC
 ACAGAGTGAGACTCATCTTAAAAAAAGTATCAGCATTCCAAACCATAGGGACAGGTGTTTTTATT
 TGTCTTCGATAATTTACTGGTGTCTAGAGGCCGAACGGGGTGCCTCTGAAAGGCACACCTCATGG
 25 GAAGAGAAATAAGTGTGAATGGTTGTTAAACAGAGGTTAAACTGGGTCCTGTGCTGAGTTAACAGTCCAGATC
 TGGACTTTGCCTTTCCAGAATGCTCCCTGGGTTGCTCATGGGGAGTGGCAGGTGCAACACCCCTGTGATGG
 GAGCAGCAGGTGCAAGCCTCATGATGGGGAGTGGCAGGTGCAACACCCCTGTGATGGCAGCAGGCTGATGG
 TTGAGCTCCCTCCCCACAAGGATGCCGCTCCGTGCTCCCACAGTCCCTGCTTCCCTCAGCCTTACCTGG
 CTGGCCTCCACTGGTTGTCTGATGTTCCACATTCCGGCTCCAGCACCTTCCGCTCTCCAGGACACCT
 30 GCAGTGCTGGCATAACAGTCAGCTGTGAACGTCCACTGCTTATTGCTCCCTGAAATGTATTAGGACAGG
 ACCCCCTGGTCCAGCCTGGCACAGCATAGTGAATGTTTGAGGACAAAGGACAGACAAACAAATCAGGAAATGG
 GTTCTCTAAACACATTGCAAAGCCACAGGGTAGTGCAGGATGGTGGCATCAGGTGATCAGATGTTGGTCAAATG
 CCAGAATATTCTGTCTCCAAAGGCCACTGGTCAAGAGTGTGCTGAGGGTGTCTAAAGCTCAGCAGTGGAG
 GCAGTGGTCCGCAACTCAGGGTAGACTCACATCTGTGCTGAAGTATACAGCAGAGGCTGAAAGGCATCTGG
 35 GAAGAAAACAGGCAAATGATTAAGAAAAGTGAAGGAAAGAGTGTGAAGATGGGATTTCTGTCAGATTTAGT
 TCCCAAACACAGCTCAGATGGTAGAATGGTCTGAGAACACTGATGGCACAGAACATAGAACAAACGGAAAGCC
 CAGAACACGAAATAAGTGTGCTTACAGCATACAGAGCAGATTCTAGGTAGAAGAGGAGACACATGCAAAC
 ACCAGCAACGAAATAAAACAAAGACTCAAAGGGAGGGAGGTGAACGTTCCCTGGGTTGGTGGGGAGGACAC
 AGGGAGGCGGATGAAACCAAGTGGAGCAGGGCATTGCTTCACTGCAAGGAAACTCAGCTGCCTGAGCCACAGT
 40 ATGGCCATTCCCTGGAGCGTTGTGACAGTGATTTATTAAAGGCCCTGTGAGGTCTGCACATTCTCATCTCACTT

5 GTTCTCTAACCACTGAGAGGTAGAGGGAGAAAGGCTCCAGGGGAGCCAGCCGCCCTGGTCACCCAGCTGGCAAGGGC
ATGCATGATTGAGCTGGCTCTGCTCGGGGCCCTGCTCTGCCAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCAAAGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAGAAATGCACTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
GAGACCTGCCCCATCCCTCATGCTGGCTCTTCTGGCTTGCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTCTGAAAGCAGCTGTTGATGGAAGTCCTCACAAATGCTCTGTCTTCCAGTAATTCAACTTCTGAAGTGA
CCAGACATTATCACGGGTCTTACCATTTCCAGTGTCCAGGCAGGGGACTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCACTCACAAACTGCTCTGCATTAAACATTTTCAAAGAATTGGAAAT
GTTTAATGGCACAAACGTTTATTCAATGTAGCAGTGTCAAAGCTGGATGTAAGAACACACCCCCAGGAGCCTGCC
10 TGAATGTCATGTGTTCATTTGGACATGGACATACATGGCAGTGAGTGGTGGTGGAGGCCCTGGAGGACATCGTGG
GATGCCCTCATCCCTGCCCTCTGGAGACACCATGTCAGTGTGCCAGTGCACTCACTGGAGCCCTGTTAGCTGGTGCACCTG
GCTCTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCAACTCAGTGTCTCCACAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGAGCCCGGGAGCCAGGGCTCCACAGTTATTATGTTGGCTGAGTTATGTCAGATC
15 TCATCAGGGCAGATGATGAGTCAGACAAACACGGCGTGGAGGTTGGATAACACTCAACATCACTAGCCAGGTCTGGT
GAGTTGGTCATGCAGAGTCGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCCAGCCCCCTGGCTGCAGC
CCATGCCCAAGGAGAACAGGAAGCGGGAGGAAGGAGGGAGCTTTGGAGCAAGCTTGCAGGAGGGGCTGGTGT
GGGGCAGGCACCTGTGTCAGCATCCCCCTGTGTCAG

Intron 12 (SEQ ID NO 16)

20	GTGAGCAGGCTGATGGTCAGCACAGAGTTCAAGAGTTCAAGGGGTGTGCGCAAGTATGTGTGTGTGCGCGCT GCCTGCAAGGCTGATGGTACTGGCTGCACGTAAGAGTGCACATGTACGCATATACACGTGAGCACATACATGTGTGCA GTGTGTACATGAAGGCATGGCAGTGTGTGCAAGGGCACAAGTGTGTGCAATGCGAATGCGCACACCTGACA TGCATGTGTTCTGTGACAGTCGTTGGGCATTACAGTGAGGTGATCGCTGTTGGGTGTGCAAGTGTGTGAGTAGCATGTG GCACATAACATGTATTGAGGGTCTCGTGTCACCCCCTAGGTCTCAGCACCAGTGCACACTCCTAACAGGATGAGAC
25	GGGGTCCCAGGCCTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGGATCCGCGTCCACT CCCTCTCTGTGGGCTCTGTGCACTCCCCCTCTGTGGGCATTACATCCACTCCACTCCCCCTCTCTGTGGG ATCCCGGTCACTCCCCCTCTGTGGGCATCTGGTCCACCTCCCCCTCTGTGGGCATTGTGCGTCACTCCCCCTCT GGTTCTTCTGTCTTGGCCAGCCTCGGGGAGGCAAGTGCACACAGTCTTGACTCGCCCAGGGTGGTTCGCAGCTG CCGGGTGAGGCCAGGCCGATTCACTGGGAAGAGGGATGTTCTGTCAAATGTTCTTCTGTGTCATCTG
30	ATGGATGATAAAGCAAAAGTAAAACCTAAATCCCAAGAGGGTTTACCGTTCTCACTCTTCTGGCAGCTGAG

Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCACCAAGGGGTGCAAGGCCAGCCTCCAGGGACCCCTCCGGCTCTGCTCACCTGTGACCCGGGCTTCACCT
TGGAACTCTGGTTTAGGGCAAGGAATGTCTTACGTTTCAGTGGTGTGCTGCTGTGACAGTTCTGTTGCGCTG
GCTCTGTGCAAAGCACCTGTTCTCCATCTGGGTAGTGGTAGGAGCCGTGTGGCCCCAGGTGTCCCCACTGTGCGCTG
GCACGGCCGTGGGACGTATGGGGCCATCCCAGGGCAGCAGGGGATGGGTAAGAGATGTTTATGGGAGTCTTAG
CAGAGGAGGCTGGGAAAGGTGTGAAACAGTAGATGGGAGATCAGATGCCGGAGGATTGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGCTGGCCCCACCCCCGGGAAGGTGCAGCAGAGCTGTGGCTCCCCACACAGCCCCCA
GCACCTGTGCTCTGGCATGGCTGTGCTCTGGAACGTTCCCTGTCTGGTCAAGGGGTGCCCAAGAATCG
40 ACAACTTTATCACAGAGGGAAAGGGCAATCTGTGGAGGGCAACAGGGCAGCTCTGCTGGATCAGGGCAGGTGGTGGC

ACAAGCCTGGGCTGTACCAAGGGCAGTCGGCACCAAGGCCGGCCTCACCTAACAGGCCCTCCAGGCCACTG
 GGAGCTGAATGCCAGGAGGCCAAGGCCCTGCCCATGAGGCTGAGAGAGGTGTGAGCATTGTGTTACCCAGGCCG
 AGGCTGCGAATTACCGTGCACACTGTGAAATGAGGCTGCTGTATCGTGGAAACCCAGCAAGGGCTCACGGGA
 GAGTTTCCATTACAAGGCGTACCATGAAAATGGTTAACCGAGTGCTTGCCTCATGCTCTGGCAGGGAGGGC
 5 AGAGCCACAGCTGCATGTTACCGCCTTGACCCAGCTCCAGGGCTTGGGACCCAGGCTGCTCAGTCCAGGGTGGTCC
 GGCTCAGACGCCCTCTCTGCCTCTCTGCCCTAAATCTCCCTGTTGCATCTCCCTGACGCCGTGCCCTGGG
 CCCTCGTGCAGCTGCTGACTCCTTCCGGAAACCCCTGGGTGTGCTGGATACAGGTCCACTGAGGACTGGAGGTG
 CTGACACTGTGGTGACCCAGGGTCCAGCTGGCGTCTGGGCTCTGGGCATGATGAGGTAGAGGAGTTTCC
 CAGGTGAAAACCTGGGAAACTCCAGGGCATGTGACCTGCCACCTGCTCCCTCCATATTCACTCAGTCTTGTCTC
 10 ATTTCCCCACCAAGGTCTCTAGCTCGAGGAGCTCCGTAGAGGGCTGGGCTCAGGGCAGGGCGGTGAGTTTCCCAC
 CCATGTGGGACCCCTGGGTAGTCGCTGATTGGTAGGCGTGGAGGAGCCGAGATGGGATGGGACACGGCCGGTTC
 AACACAGAGTCAGGCACGTGAAGGCCAGGAATCCCTCCCTGAGGCAGGGAGTGGAGAACGGAGAGCTGGCCCG
 ATTTCAGGCAGGCAGGTGCACTGGGAGGCTGTGGTGGTCCACGTGGCGTGGGGCGGGGCTGATTCAAATCCG
 15 TGGGCTGGGCTTCCCTGGCCGTCTGGCCGCGCTCACACGGCTGGGGTGACGCCCGACCTCTAGCAGGTGGC
 CCGGGACCTTAGGCTTATTATTTGTTAAAACATTCTGGGCTGGCTTCCGGTGTGCTAAATGGGAAAAGACATCC
 CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTGCTGGCTTGACTGGTGATCTCAGGTCACTCAGAAGTGGC
 CAGGAAGTCAGTGGAGACCAAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGT
 GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGTGACACCAGGTAC
 20 ACGGGGGCTCTGATCACACGACATATGAGCACATGTGCACATGTGCTTTATGGTAGCCAGGTCTGTGCACACCTG
 CCCAAAGTCCAGGAAGTCAGGAGGCCAAGATGGGGCTGACAGGGCTGGCGCGTGGCTCACACCTGAGTCCAGCA
 CTTTGGGAGGCCAGGCAGGAGGATCCCTGAGCCAGGAGTTAACGACAGCCTGAGCAACATAGTAGAACCCATCTC
 TATGAAAAATAAAACAAAAATTAGCTGAACATGGTGGTGTGCGCTGTAGTCCAATACTGGGAGGCTGAAGTGGAG
 GATCACTTGAGCCCAGGAGGTGAAGCTCAGTGAGCTGAGATTGACCAACTGTACTGCGACCTGGTGACAGACTGAGA
 25 GCCCATCTCAACAACAACAAAGAAGACTGACAAATGAGCTTCTGGAAAGAAACATTAGTAGGAACCTAACCTACACA
 CAGAAGCCAAGTCGGTGTCTGGTGTAGTGAGATGGGTCTCACACCATCACCCAGACCCAGGGTTATG
 CACCAAGGGCGGGTGGCTCAGAAGGGATGCCAGGACGGTGTGATATACGATGACATCAAGGTTGCTGACGAAGGGCAG
 GATTCACTGATAAGTACCTGTTACAAAGGAACAATGGATAACTGAAACCTTAGGGCTTCCCGAACAGGGCT
 AATCAGAAGGCCAGCATGGGGGCTGGCATCCAGGATGGAGCTGCTCAGCCTCCACATGCGTGTACAGATGGTGA
 30 CAGAAACCGAGTGTACCTGTGCACACACAGACACCGAGCTACTCGCACACACAAGCACACACAGACATGCGATG
 ATCCGTGTGTGCACCTGTGCCATGAGGAAACCCATGCGATGTGCATTGACCGCACACAGGCCGGTGGCCCAT
 GCCCACACCCACGAGCACCGTGTAGGAGGCCCTTCCCTGACGCTGTCCGCCATCCCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCCTGGCCTCAGTGGCAGCAGTCGCTGCTGGTTAGTGTGTCAGGAGACTGAGTGAATCTGGG
 CTTAGGAAGTCTTACCCCTTTCGCATCAGGAAGTGGTTAACCCACCAACTGTCAAGCTGCTGCCCCCTCTCGT
 GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGCTGGGAGCTGCCATCCTCCACCTTGTCTGCTGGGAA
 GCGCTGGGGGCTGGTCTCCTGTTGGCCATGGTGGATTGGGGGCTGGCTCTCCTGTTGCCCTGTGGTGG
 GATTGGGCTGCTCTCCGTCATGGCACTTGGCACTTGGGCCCTGTGCAAACCCAGGCCAGGGCTTGGAGGAGGCCAG
 40 GCTACCCACCCCTCTCAGGAGCAGGCCGTATCACACAGACAGGCCGCCGTCTGCTTCCCAGTCACCG

TCCTCTGCCCTGGACACTTGTCCAGCATCAGGGAGGTTCTGATCCGCTGAAATTCAAGCCATGCGAACCTGCGGT
 CCTGAGCTAACAGCTTCACTTCTGTTCTTCTGTGTGAGGAAATTCAACCTGGAGAAGCCGAAGAAAACATTCTG
 TCGTGACTCCTGCGGTGCTGGCTGGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGAGCTTCCG
 GTGTCTCCTGGGAGGGAGCTGGCTGGCCTGTGACTCCTCAGCCTCTGTTTCCCCAG

5

Intron 15 (SEQ ID NO 19)

GCAAGTGTGGTGGAGGCCAGTCGGGCCCCACCTGCCAGGGTCATCCTGAACGCCCTGTGTGGCGAGCAGCCTC
 AGATGCTGCTGAAGTGCAGACGCCCGGGCTGACCCCTGGGGCCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
 CTGGTGTCCCAGGCCAGGCTGGCAGGGTCCCAACTCTTGAACCCCTGCTTCCATCTCAGGGCGATGGCTCC
 10 CCACGCTTGGGAGCCCTCTGACCCCTGACCTGTGTCCTCTCACAGCCTTCTGGCTGCTGCCCTGAGCTCCTGGGT
 CCTGAGCAAGTTCTCTCCCGCCCGCTCCAGCCTGACTGGCTGCTGTGCTGCCCGGTGGAGGGGTGCTG
 TCCCTCACTGAGGTTCCCACCAGCCAGGCCACAGGGTGCAGGCCCTGCCTGCCGCCACCCACAGCTCTAGGAGGG
 TTGGAGGATGCCACCTCTGCCCTCTGGAACGGAGTCTGATTTGCCCGCAG

15

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATTTGAATCTTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAACGCTGTGAGTGAACGGGTGGT
 GGTCACTGCGGGCCCATGGCTGGCTGTGCAATTACGAAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCAGGCCCCATG
 GCCTGGCTGGGCTGGAGGTTCTGATGCTGTGAGGCCAGGAGGGAGGGTAGGGATAGACAGTGGGAGCCCCA
 CCCTGGAAGACATAACAGTAAGTCCAGGCCAGGGCAGCAGGGATGCTGGGGCCAGCTTGGCGGGATGATG
 20 GAGGGCCTGCCAGGGTGGCAGGGATGATGGGGCCAGCTGGGGTGGCAGGGGTGATGGGGGGCTGGTCTGGTGG
 CGGGGAAGATGGGAGGCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
 GGTGCACATCTCTGGGCCATCAGCTTCACTGGAGGTGGGGCAGGGCATGACACCATCCTGTATAAAATCCAGGATT
 CCTCTCTGAACGCCCAACTCAGGTGAAGTCACATCCGCTCTGGCATTCTTAAGAGTAGACCAAGGATTCTG
 ATCTCTGAAGGGTGGTAGGGTGGGCACTGGAGGTGTGGACACAGGAGGCTCAGGGTGGGGCTGGTGTGCTCTC
 25 ATCCTCTTATCATCTCCAGTCATCTCATCTCTTATCATCTCCAGTCATCTGTCTTCTCTTATCTCCAGT
 CTCATCTGTCATCTCTTACCATCTCCAGTCATCTCTTATCTCTTATCTCTAGTCATCCAGACTTACCTCCA
 GGGCGGGTCCAGGCTCGCAGTGGAGCTGGACATACGTCCTCCCTCAGCAGAAGGAACGGAAAGGATTGCAAGAACAG
 GAGGGGCGGCTCAGAGGACGCACTGGGGTGAAGAACAGCCCTCTCAGAAGTTGGCTGGGCCACAGAACCG
 AGGGCCCTGCGTGAAGTGGCTCCAGAGCCTCCAGCAGGTCCCTGGTGGGGCTTATGGTATGGCCGGTCTACTGAGTG
 30 CACCTGGACAGGGCTCTGGTTGAGTCAGCCGGACGTGCTGGTGGGGTGGGGCTTATGGCACTGGATATG
 GCGTCATTTATTGCTGCTGCTTCAGAGAATGTCAGTGAACGGAGCCTAATGTTATGGTGGGGCCAAGTCCACAGACTG
 TGCGTAAATGCACTCTGGTGCCTGGAGGCCCTGATAGGAGCTGTGAGGAAGGAGGGCTTGGCAGCCGGCTGGG
 GCGCTTTGGCCCTGCAAACCTGGAGGGAGCGGCCCGGGCGCCGTGGCGACGACCTCAAGTGAAGAGGTTGGACAGAAC
 AGGGCGGGACTTCCAGGAGCAGAGGCCCTGCTCAGGCACACCTGGTTGAATCACAGACCAACaGGTCAGGCCATT
 35 GTTCAGCTATCCATCTTACAAAGCTCCAGATTCTGTTCTCGGGTGTGTTGGTGAATTTTACTCAGGATTACT
 TATATTGCTAAAGTATTAGACCTAAAAAGGTATTGCTTGTATGGCTTAACACTCAAGCACCTACTTAT
 TTGCTGTTTATTATTATTATTATTAGAGATGGTGTACTCTGTCACCCAGGTTGTTAGTCAGTGGCAG
 AGTCATGGCTCGCTGTAGCCGAAACCCCAAGGCTCAAGTGAATCCTCCGGCCTCAGCTCCAGAGTGCTGGATTACAG
 GTGTGAGGCCACTGCCCTGGCAGTAAACCAACTATGAAAGTCAGGTCCAGTGGCTTACACCTGTATCC
 40 CAGTAGTTGGAGGCCAGGAGCAGAAGGATTGTCAGGCCAGGAGTTGAGACCAAGCATGGGTAACATAGGAGACCC

ATCTCTACAAAAAATGCAAAAGTTATCCGGCGTGGGTCCAGCATCTGTAGTCCAGCTGCTCGGAGGCTGAGTGGG
AGGATCGCTTGAGCCGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCAGTCCAGCCTGGCAACAGAGTGA
GACCCCTGTCTCAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGAAGGAAAGAAGAAGAAGAAGAAGAAG
GAAGAAGGAAGAAGAAGGAGAAGGAGGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAATGAAAATAACA
5 AAGTTTAAAGGGAAAGAAAAACCCAGCTTTGGACTTCCTTAGGCTGAACCTCATCTCAAGCAGCTTCCTTCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGCACCTGCTGTAACCGTC
GATGTTGGTGCAGGTGCCACCTGGGAAAGGATGCTGTGCGAGGGGCTGCCAAACTTGGGGTTTCAGAAGCCCCAG
GCCACTTGTGGCAGGCACAATTACAGCCCCCCCCAAAGATGCCACGTCCTTCTCCCTGGAACCTGTGAATGTGTACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCACTCTGGATTATCTGG
TGGGCTGTATGGCACAAAGGTCCCTAGAAGTGAGAGAGGGAGGAGGGAGACTCAGAGAGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTGAGATGGAGGAGGGGTCCCAGCCAAGGAATGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCGGTCTGAGGGCACACGGCCCTGCCACGCCCTGATTCAGGCCAGTGGGACCTGTTCAGCTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTGTGTTCAAGGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAATGAAATAG
15 CAGTACAGGGAAATGAATAACAGGGACAGTTCTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-
5 charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative 15 Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 20 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des 25 offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von 30 Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

5 Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, lässt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

10 Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer interner 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

15 Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

20 25 Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind.

30 Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur 10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv markierter Primer (5'GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵ cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL), 2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl 15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden 2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150µl 20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5 min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel 25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom 30 ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA_nTTGT) repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzutragen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den 10 Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgenen (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.
20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem 25 genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 30 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-5', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTCCTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 bis -171; Primer RP-3A: 5'-10 CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

20 In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. 25 bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

30 Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschiedene lange hTC-Promotorsequenzabschnitte generiert und 5' vor 10 das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 15 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung 20 bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-25 CCAGATCTCTGGAACACAGAGTGGCAGTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaars PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTGCAG-3', Position -3120 bis 3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmente 5 (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. 25 CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszuplattieren.

10

15 Für die Transfektion wurden pro Well jeweils 2x 10⁵ -HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serumfreies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl 20

25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₂; Mg(OH)₂·H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysat und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na₂CO₃ gestoppt und die Absorption bei 420 nm bestimmt.

5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK 10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen 15 Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

5 Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Furcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.

10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Whiley-Intersciences, New York.

15 Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.

20 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.

25 Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.

30 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.

35 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.

40 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.

45 Harley, C.B., Furcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.

5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.

10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.

15 Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.

15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.

20 Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.

20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.

25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.

30 Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.

30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.

35 Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.

35 Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 **Shay, J. W. (1997).** Telomerase and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

10 **Vaziri, H., Dragowska, W., Alisopp, R. C., Thomas, T. E., Harley, C.B. und Lansdorp, P.M. (1994).** Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

15 **Zakian, V. A. (1995).** Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.

5

2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.

10

3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.

15

4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.

20

5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.

6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.

25

7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.

8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

30

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:

5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.

15 11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.

20 12. Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

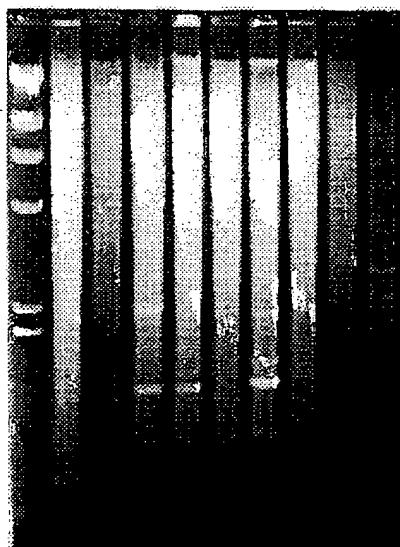
25 A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält mit Körperflüssigkeiten oder zellulären Proben,

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1**A**

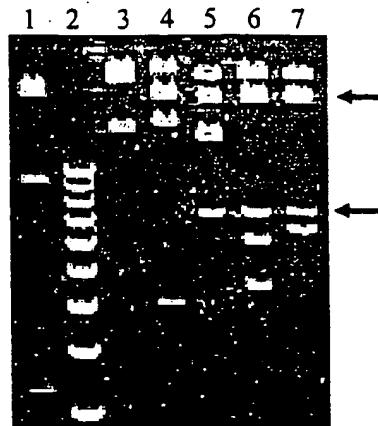
1 2 3 4 5 6 7 8 9 10

**B**

1 2 3 4 5 6 7 8 9 10



Fig. 2



ERSATZBLATT (REGEL 26)

Fig. 3

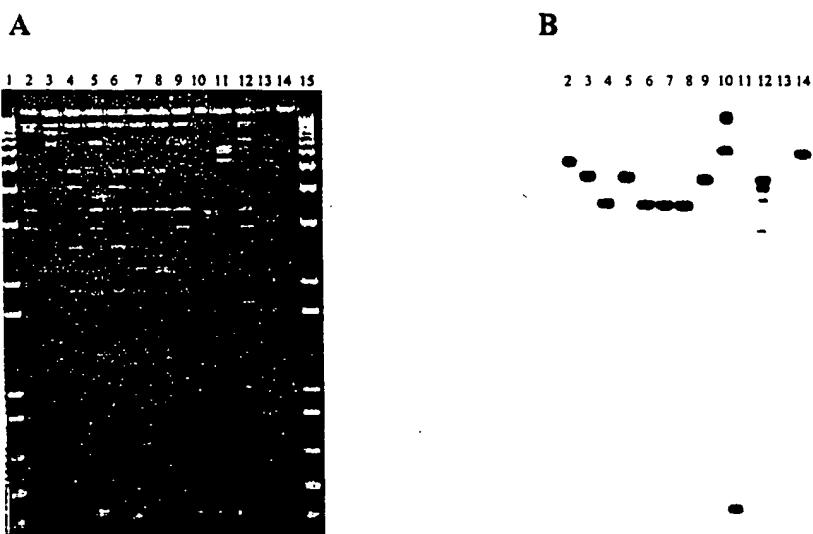


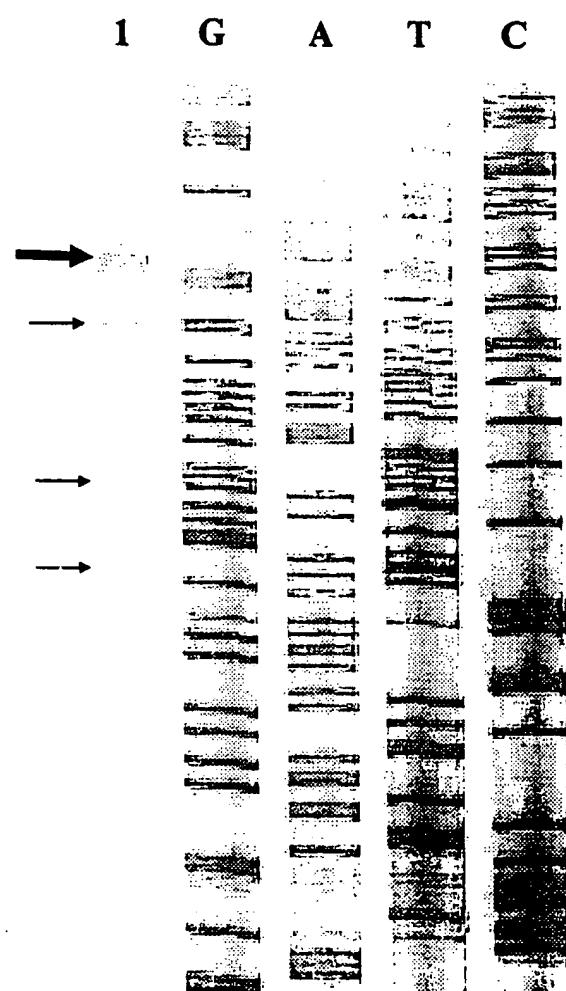
Fig. 4

GAGCTCTGAA CCGTGGAAAC GAACATGACC CTTGCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA 70
 GTGGTGTGCA GGAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTCCCTTC CATCATTATT 140
 CATTTTACCC CCCAAGGACT GAATGATTC AGAACCTTC TCGGGTGTGA CAAGCCATGA CAAAACCTCAG 210
 TACAAACACC ACTCTTTAC TAGGCCACA GAGCACGGGC CACACCCCTG ATATATTAAG AGTCCAGGAG 280
 AGATGAGGCT GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCAGGCTGAAAC AGTCTGTTCC TCTAGACTAG 350
 TAGACCTGG CAGGCACTCC CCCAAATCT AGGGCCTGGT TGCTGCTCC CGAGGGGCCG ATCTGCCCTG 420
 GAGACTCAGC CTGGGGTGC ACACGTGAGGC CAGGCCCTGTC TCCACACCC CCGCCCTCCAG GCCTCAGCTT 490
 CTCCAGCAGC TTCTTAAACC CTGGGTGGGG CGTGTTCCAG CGCTACTGTC TCACCTGTC CACTGTGTCT 560
 TGTCTCAGCG ACCTAGCTCG CACGGTCTCT CTCACATGG GGTGCTGTC TCCTTCCCTC ACACTCACAT 630
 CGCTGTGAGG GAGGAGATTC TGCCGCTCCC AGACTGGCTC CTCTGAGCC CAACTGGCT CGTGGCCCCC 700
 GATGAGGTT CCTGGCGTCC GGCTGCAACCC TGACCTCCAT TTCCAGGCC CCCCCCTCTC CTGTCATCTG 770
 CGGGGGCTG CGGGTGTGTT TTCTGTTTC TGCTGCTCCCT TCCACGTCCA GCTGCGTGT TCTCTGCCCG 840
 CTAGGGTCTC GGGGTTTTTA TAGGCATAGG ACGGGGCGGT GGTGGGCCAG GGCCTCTTG GGAAATGCAA 910
 CATTGGGTG TAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGCA CAGGGCTGGG GATGGAGCCC 980
 CGCCAGGGG CCCGCCCCCCTC ACTTCTCTG CCCCCCTCCCT CTGGAACACA GAGTGGCAGT 1050
 TTCCACAAGC ACTAACGATC CTCTTCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCCACAGC 1120
 CCTGGGAAATT CACGTGACTA CGCACATCAT GTACACACTC CGCTCCACGA CGCACCCCCC CTGTTTATT 1190
 TAAATAGCTA CAAAGCAGGG AATCCCTGC TAAATGTC TAAACAAAC TGTTAAACA AACGGGTCCA 1260
 TCCGACCGT GGAACAGTTC TCACAGTGA GAGGACATG CGCTTATAA AGCTTCAGG CATCTCAAGG 1330
 GAATTACGGT GAGTCAAAAC TGCCACCTGC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC 1400
 ACCCCATGGC AGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCCCA GCTGGGGCT ACTGCACCCA 1470
 CCTTTTACTA AAGCCAGTT CCTGGTTCTG ATGGTATTGG CTCACTTATG GGAGACTAAC CATAGGGAG 1540
 TGCGGATGGG GGAACCCGGA CGCTGTGCCA TCTTGCAT GCGGAGTGT CCTGGCAGG ATAATGCTCT 1610
 AGAGATGCCG ACCTCTGAT TCCCCAACAC CTGTCAGCAG AACCCCGCCG GCCCCAGGGC CTTGCAAGGT 1680
 GTATCTCCG TGAGGACCCCT GAGGTCTGGG ATCTTCGGG ACTACCTGCA GCCCCGAAA GTAATCCAGG 1750
 GCTTCTGGG AGAGGGGGC AGGAGGGTCA GAGGGGGCCG CCCTCAGGAC GATGGAGGCC GTCACTGTA 1820
 GGCTGAAAAG GGAGGGAGGG CCTCGAGCCC AGGCTGCCA GCGCCCTCAG AAGCTGAAA AAGGGGAA 1890
 GGGACCTCC ACAGGACCTG CACCGAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC 1960
 TCCGGCTCC GTGCCATAGG AGGGCCTGCG CGCTGCCCTT CTAGCATGAA GTGTTGGGG ATTGCAAGAA 2030
 GCAACAGGAA ACCATGACAC TGTGAATCTA GGATTATTTC AAAACAAAGG TTACAGAAA CATCCAAGGA 2100
 CAGGGCTGAA GTGCCCTCCGC GCAAGGGCAG GGCAGGCACG AGTGAATTAA TTACAGTATT TTATTTATT 2170
 TACTTACTTT CTGAGACAGA GTTATGCTC TGTGCCCCAG GCTGGACTGC AGCGCATGAA TCTTGGCTCA 2240
 CTGCAACCTC CGTCTCTGG GTTCAAGCA TTCTCGTCC TCAGCCTCCC AAGTACTGTT GATTTCAGGC 2310
 GTGCCACCAC ACACCCGGCT AATTTGTTAT TTTTACTGAA GATGGGTTT CACCTGTTG GTCAAGCTGAA 2380
 TCTCRAAATC CTGACCTCAG GTGATCTGGC CACCTCAGG TCCCAAAGT CTGGGATTAC AGGCATGAGC 2450
 CACTGCACCT GGCCATTATTA ACCATTTTAA AACTCCCTG CCGCTCAAGC ACACCAACTG GTAAGGAGTT 2520
 CATGGAGTTC AATTTCCCTT TTACTCAGGA GTTACCTCC TTTGATAATT TCTGTATTTC TTCTGAGACT 2590
 GGGGATAACAC CGTCTCTGA CATATTCA CTTCTGTGA CCACCTGTTA TCCCAGGGG CCCACTGCA 2660
 GGGCAGCTGG GGGGCTGGAG CGCTTCAGGTC CCAGTGGGGT TGCCCATCTGC CAGTAAACAC CTGATGTAGA 2730
 ATCAGGGCAGC AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCACTGTTG GCTGAAACAT GTAGAAATTAA 2800
 AAGTCCATCC CTCTTACTCT ACTGGGATTG AGCCCCCTCC CTACCCCCC CCAGGGCAG AGGAGTTCT 2870
 CTCACCTCTG TGAGGAGG AATGATACTT GTTATTTTT CACTGCTGGT ACTGAACTCA CTGTTTCTATT 2940
 TGTGGTTTG TTGTTTTGT TTGAGAGG CGTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTCAATGG 3010
 CGCAGCTTG GCTTACTGCA GCCTCTGCC CCGAGGTCA AGTGATTCTC CTGTTCTCCGC CTCCCATTG 3080
 GCTGGGATA CAGGCACCC CGACCATGCC CAGCTTAATT TTGTTATT TAGTAGAGAC GGGGGTGGGT 3150
 GGGGGTCACT ATGTTGGCCA GGCTGGCTC GAACTCTGA CCTCAGATGAA CCTACCTGCC TCTGCCCTCCT 3220
 AAAGTGTGG GATTACAGGT GTGAGGCCAC ATGCCACTG CAGAATTAC TCTGTTAGA AACATCTGG 3290
 TCTGAGGTAG GAGGCTCACC CCACTCAAGT GTTGGGTGT TTAAGCCAA TGATGAAATT TTTTATTGT 3360
 TGTAGAACAA CTCTTGAATGT TTGACTCTG AGTGACTAAG ACATCATGAG CTTTCAAG ACACACTAAC 3430
 TGCACCCATA ATACTGGGAT GTCTCTGGG TATCAGGAT CTTCATGAA TGCCGGAGG CGTTTCTCG 3500
 CCATGACACAT GTGTTTAATT ACTCCAGCAI AATCTCTGC TTCTTATTCT CTCTTCCCT CTTTIAAAAT 3570
 TGTGTTTCT ATGTTGGCTT CTCTGAGAG ACCAGTGTAA AGCTACAACT TAACTTTGT TGGACAAAT 3640
 TTCCAAACC GCCCCCTTGC CCTAGTGGCA GAGACAATTC ACACACACAG CCCTTAAAAA AGGCTTAGGG 3710
 ATACATTAAGG GGATTCTAG AAGACCGACC TGTAATCTA AGTATTACAG AGACAGGGCT AACCTCCAGC 3780
 GAGCGTACA CAGGCAGGGAG GGTGGAGGC CGTCTGAAAGT CTGAGCTCCA TAATAAAGC AATTTCTCC 3850
 GGCAGTTCT GAAAGTAGGA AAGGTTACAT TTAAAGGTGCG TTGTTGTAGC ATTCAGTGT TTGCCCAGCT 3920
 CAGCTACAGC ATCCCTGCCA GGCCTCGGG AGCCAGAGG TTCTCGCCC CCTTAGATCC AACATTGAGC 3990
 AACCCGGAGT CTGGATTCTC GGGAGTCTC CAGCTGTCTC GCGGTTGTGC CGGGGCCCA GGTCTGGAGG 4060
 GGACCCAGTGG CGCTGTGGG TCTACTGTC CGCTGGAAAGT CGGGCCCTCC AGCTCTGCG TCCGAGGGCTT 4130
 GGAGCCAGGT GCTCTGCCAC CGAGGCTGCC CTCCACCTG TGCGGGGGGG ATGTGACCGAG ATGTTGGCCT 4200
 CATCTCCAG ACAGAGTGTCC GGGGCCAGG GTCAAGGGCTG TTGTTGGCTGG TGTTGGCCG CCGGTGCGCG 4270
 GCCACCGAGA CGCCCTGGCT CCATTTCCCA CCCTTCTCG ACGGGACCCG CCGGGTGGGT GATTAACAGA 4340
 TTTGGGGTGG TTGCTCATG GTGGGGACCC CTCGCCGCT GAGAACCTGC AAAGAGAAAT GACGGGCCCTG 4410
 TGTCAAGGAG CCGAAGTGC GGGGAAGTGT TGCAAGGGAGG CACTCCGGGA GGTCCCCCGT GCGGGTCCAG 4480
 GGAGCAATGC GTCTCGGGT TGCTCCCTAG CGCGCTCTAC CGCCTCCCGT CCTCCCTTC ACCTCCGGCA 4550
 TTCTGTTGTC CGGGAGGCC ACGGCCCCG CTCGGACCTG GAGGCAAGGCC TGGGGTCTCG GATCAGGCCA 4620
 CGGGCCAAAG GTGCGCCGCA CGCACCTGTT CCCAGGGCT CCACATCATG GCCCCCTCCCT CGGGTTACCC 4690

Fig. 4 (Fortsetzung)

CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGGTC CCTGCACCCCT GGGAGGCCGA 4760
GGGGCGCGCGG GGCGGGGAAG CGCGGCCAG ACCCCCCGGT CGGCCCGGAG CAGCTCGCT GTCGGGGCCA 4830
GGCGGGGCTC CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCGGGTCCC 4970
GACCCCTCCC GGTCGGCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCC CCCCTTCCCTT TCCGGGCC 5040
CGCCCTCTCC TCGCGGCCCG AGTTTCAGGC AGCGCTGCCT CCTGCTGCC ACGTGGGAAG CCCTGGGCC 5110
GCCAACCCCCC GGGATG 5126

Fig. 5



ERSATZBLATT (REGEL 26)

Fig. 6

GGTTCAAGGCA	GCGCTGGTCA	CTGCTGGCA	CGTGGGAAGC	CCTGGCCCCG	GCCACCCCCG	CGATGCCGCG	70
CGCTCCCCCG	TGCCGAGGCC	TGCGCTCCCT	GCTGGCAGC	CACTACCGCG	AGGTGCTGCG	GCTGGCCACG	140
TTCGTCGGC	GCTGGGGCC	CCAGGGCTGG	CGGCTGGTC	AGCGCGGGGA	CCCGCGGGCT	TTCCCGCGCG	210
TGGTGGCCCA	GTGCTGGTGC	TGGTGGCCCT	GGGACGCACG	GCCGCCCCCC	GCCGCCCCCT	CCTTCCCCCA	280
GGTGTCTGC	CTGAAGGAGC	TGGTGGCCCG	AGTGTGCGAG	AGGCTGTGCG	AGCGCGGGCG	GAAGAACGTG	350
CTGGGCTTCG	GCTTCGCGCT	GCTGGACGGG	CCCCGGGGGG	GCCCCCCCCG	GGCCTTCACC	ACCAGCGTGC	420
GCAGCTACTT	GCCCCAACACG	GTGACCGACG	CACTGGGGGG	AGCGGGGGG	TGGGGGCTGCG	TGCTGGCCCG	490
CGTGGCGAC	GACGTGCTGG	TTCACCTGCT	GGCACGCTGC	GCGCTCTTTC	TGCTGGTGGC	TCCCAGCTGC	560
GCTTACCAAGG	TGTGCGGGCC	CCGGCTGTAC	CAGCTGGCG	CTGCCACTCA	GGCCCGCCCC	CCGCCAACACG	630
CTAGTGGACC	CCGAAGGCGT	CTGGGATGCC	AAAGGGCCCTG	GAACCATAGC	GTCAAGGGAGG	CCGGGGTCCC	700
CCTGGGCTG	CCAGCCCCGG	GTGCGAGGAG	GGCGGGGGGC	AGTGCCAGCC	GAAGTCTGCC	GTTGCCRAG	770
AGGCCCAAGGC	GTGGCGCTGC	CCCTGAGCGG	GAGCGGACGC	CCGTTGGCCA	GGGGTCTGG	GCCACCCGG	840
GCAGGACCGG	TGACCGAGT	GACCGTGGTT	TCTGTGTGTT	GTCACCTGCC	AGACCCGGCG	AAAGAACCCAC	910
CTCTTGGGAG	GGTGCCTGCG	CTGGCACCGC	CCACTCCAC	CCATCGCTGG	GGCCGACGCA	CCACCGGGCG	980
CCCCCATCCA	CATCGGGGCC	ACACAGTCCC	TGGGACACGC	CTTGTCCCCC	GGTGTACGCC	GAGACCAAGC	1050
ACTTCTCTA	CTCCCTAGGC	GACAAGGAGC	AGCTGGGGCC	CTCTTCTTA	CTCAGCTCTC	TGAGGCCCAG	1120
CCTGACTGGC	GCTCGGAGGC	TCTGGAGAC	CATCTTCTG	GTTTCAAGC	CCTGATGCC	AGGGACTCCC	1190
CGCAGGTTGC	CCCGCCTGCC	CCAGCGCTAC	TGGCAATAGC	GGCCCCCTTT	TCTGGAGCTG	CTTGGGRACCC	1260
ACGGCGAGTG	CCCCTACGGG	GTGCTCTCTA	AGACGCACTG	CCCCTGGCGA	GCTGGGCTCA	CCCCAGCAGC	1330
CGGTGTCTG	GGCCGGGAGA	AGCCCCAGGG	CTCTGTGGCG	GGCCCCGGAG	AGGAGCACAC	AGACCCCCGT	1400
CGCCTGGTGC	AGCTGCTCCG	CCAGCACAGC	AGCCCCCTGGC	AGGTGTACGG	CTTCGTCGGG	GCCTGCGCTGC	1470
GCCCCCTGGT	CCCCCCAGGC	CTCTGGGGCT	CCAGGCACAA	CGAACGGCC	TTCCTCAGGA	ACACCAAGAA	1540
GTTCATCTCC	CTGGGGAGAC	ATGCCAACGT	CTCGCTGCAG	GAGCTGACCT	GGAAAGATGAG	CTGGGGGAC	1610
TGGCCTGGC	TGCGCAGGAG	CCCAGGGGTT	GGCTGTGTT	GGCCCGCAGA	GCACCGCTCG	CTGTGAGGAGA	1680
TCCCTGGCAA	GTTCCTGCAC	TGGCTGATGA	TGTGTACGT	CTGCGAGCTG	CTCACGGCTT	TCTTTATGT	1750
CACCGAGACC	ACGTTTCAA	AGAACAGGGT	CTTTTCTAC	CGGAAGAGTG	TCTGGAGCAA	GTTGCAAAGC	1820
ATTGGAATCA	GACAGCACTT	GAAGGGGTG	CAGCTGGGGG	AGCTGTCGGA	AGCAGAGGTC	AGCCAGCAC	1890
GGGAAAGCAG	CCCCCCTCG	CTGACGTCGA	GACTCCCCCT	CATCCCCAAG	CCTGACGGGC	TGCGGCGGAT	1960
TGTGAACATG	GACTACGTCG	TGGGAGCCAG	AACGGCCATC	CTCTACCTTG	ACAGACCTCC	TCTCACCTCG	2030
AGGGTGAAGG	CACTGTTCA	CGTGCCTCAAC	TACGAGGGGG	CGGGCGGCC	CGGCCCTCTG	GGGCCCTCTG	2100
TGCTGGGCCT	GGACGATATC	CACAGGGCCT	GGCGCACCTT	CGTGCTGCGT	GTGGGGGCC	AGGACCCGCC	2170
GCCTGAGCTG	TACTTGTCA	AGGTGGATGT	GACGGGGCG	TACGACACCA	TCCCCCAGGA	CAGGCTCAGC	2240
GAGGTCTATCG	CCACGATCAT	CAAACCCCCAG	AAACAGCTACT	CGCTGGCTCG	GTATGCCGTG	GTCCAGAGG	2310
CCGGCCATGG	GCACGTCGCC	AAGGCCCTCA	AGAGCCCCATG	CTCTACCTTG	ACAGACCTCC	ASCCGTACAT	2380
GCQACAGTTC	GTGGCTCA	TGCAAGGAGC	CAGGGCGCTG	AGGGATGGCG	TCGTCATCGA	CGACAGCTCC	2450
TCCCTGAATG	AGGCCAGCAG	TGGCTCTTC	GAACGCTTC	TACGCTTCAT	GTGCCACAC	GGCGTGCAGCA	2520
TCAGGGCAA	GTCTTACGTC	CACTGCCAGG	GGATCCCCA	GGGCTCCATC	CTCTCCACG	TGCTCTGCAG	2590
CCTGTCTTAC	GGCGACATGG	AGAACAGGAT	GTGGGGGG	ATTGGGGGG	ACGGGGCTGCT	CTTGGCTGGAG	2660
GTGGATGATT	TCTTGTGGT	GACACCTCA	CTCACCCACG	CGAAAAACCTT	CCTCAGGAC	CTGGTCCCGAG	2730
GTGTCCCTGA	GTATGGCTGC	GTGGTAACT	TGGGAAAGAC	AGTGGTGAAC	TTCCTCTGAG	AAACAGGAGG	2800
CCTGGGTTGC	ACGGCTTTG	TTCAGATGCC	GGCCCACGGC	CTATTCCCT	GGTGGGGCCT	GCTGCTGGAT	2870
ACCCGGACCC	TGGAGGTGCA	GAGCGACTAC	TCCAGCTATG	CCCCGACCTC	CATCAGACCC	AGTCTCACCT	2940
TCAACCCGG	CTTCAGGGT	GGGAGGAACA	TGCGTCGCAA	ACTCTTTGGG	GTCTGGCC	TGAAGTGTCA	3010
CAGGGTTTT	CTGGGATTTGC	AGGTGAACAG	CCTCCAGACG	GTGTGACCA	ACATCTACAA	GATCCTCCTG	3080
CTGCAGGGCT	ACAGGTTTCA	CGCATGTTG	CTGCAGCTCA	CATTCTACCA	CGAACATTGG	AAAGACCCCC	3150
CATTTTCT	GGCGCTCATC	TCTGACACGG	CTCTCCCTG	CTACTCCATC	CTGAAGCCA	AGAACGCAGG	3220
GATGTCGCTG	GGGGCCAAGG	GGCGCCGGG	CCCTCTGCC	TCCGAGGCCG	TGCAGTGGCT	GTGCCACCAA	3290
GCATTCCTGC	TCAAGCTGAC	TCGACACCGT	GTCACCTACG	TGCCACTCT	GGGGTCACTC	AGGACAGGCC	3360
AGACGGCAGCT	GATGCGGAAG	CTCGGGGGGA	CGAGGCTGAC	TGCCCTGGAG	GGCCGAGCCA	ACCCGGCACT	3430
GCCCTCAGAC	TCAAGACCA	TCTGGACTG	ATGGCCACCC	GCCCCACGCC	AGGGCCAGAG	CAGACACCCAG	3500
CAGCCCTGTC	ACCCCCGGCT	CTACGTCACA	GGGAGGGAGG	GGGGCCAC	ACCCAGCCCC	GCACCGCTGG	3570
GAGTCTGAGG	CCTGAGTGAG	TGTTTGGCCG	AGGCCTGAT	GTCCGGCTGA	AGGCTGAGTG	TCCGGCTGAG	3640
GCCTGAGCGA	GTGTCCAGCC	AAGGGCTGAG	TGTCAGGAC	ACCTGCGCTC	TTCACCTCCC	CACAGGCTGG	3710
CGCTCGGCTC	CACCCCCAGGG	CCAGCTTTTC	CTCACCAAGGA	GCCCCGCTTC	CACTCCCCAC	ATAGGAATAG	3780
TCCATCCCCA	GATTGGCCAT	TGTTCA	GGGGGATTTG	CTCCCTTGCC	TTCCACCCCC	ACCATCCAGG	3850
TGGAGACCCCT	GAGAAGGAGCC	CTGGGAGCTC	TEGGAATTG	GAGTGRACAA	AGGTGTGCC	TGTACACAGG	3920
CGAGGACCCCT	GCACCTGGAT	GGGGGTCCCT	GTGGGTCAA	TTGGGGGAG	GTGCTGTGGG	AGTAAAATAC	3990
TGAATATATG	AGTTTTCA	TTTGAAAAA	AAAAAAA	AAAAAAA	AA		4042

Fig. 7

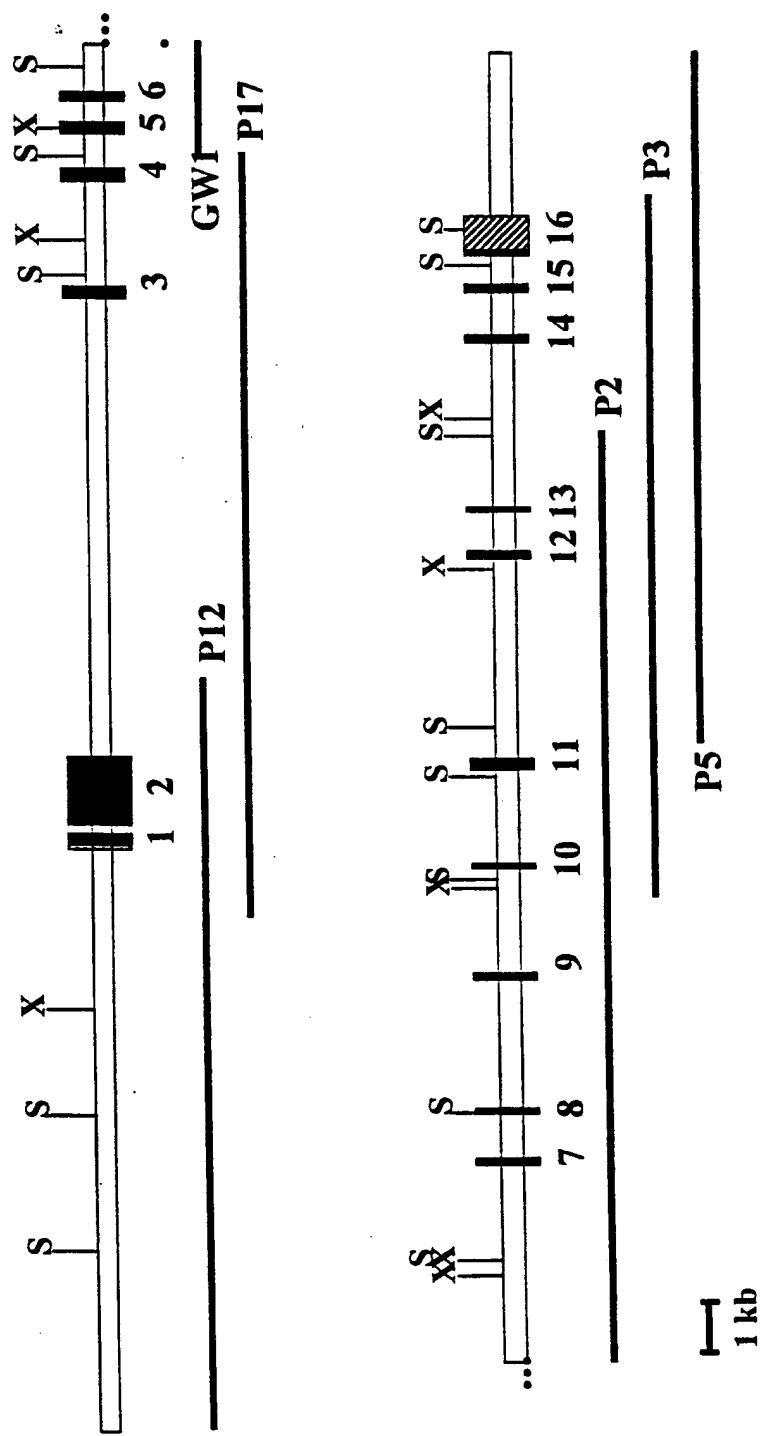


Fig. 8A

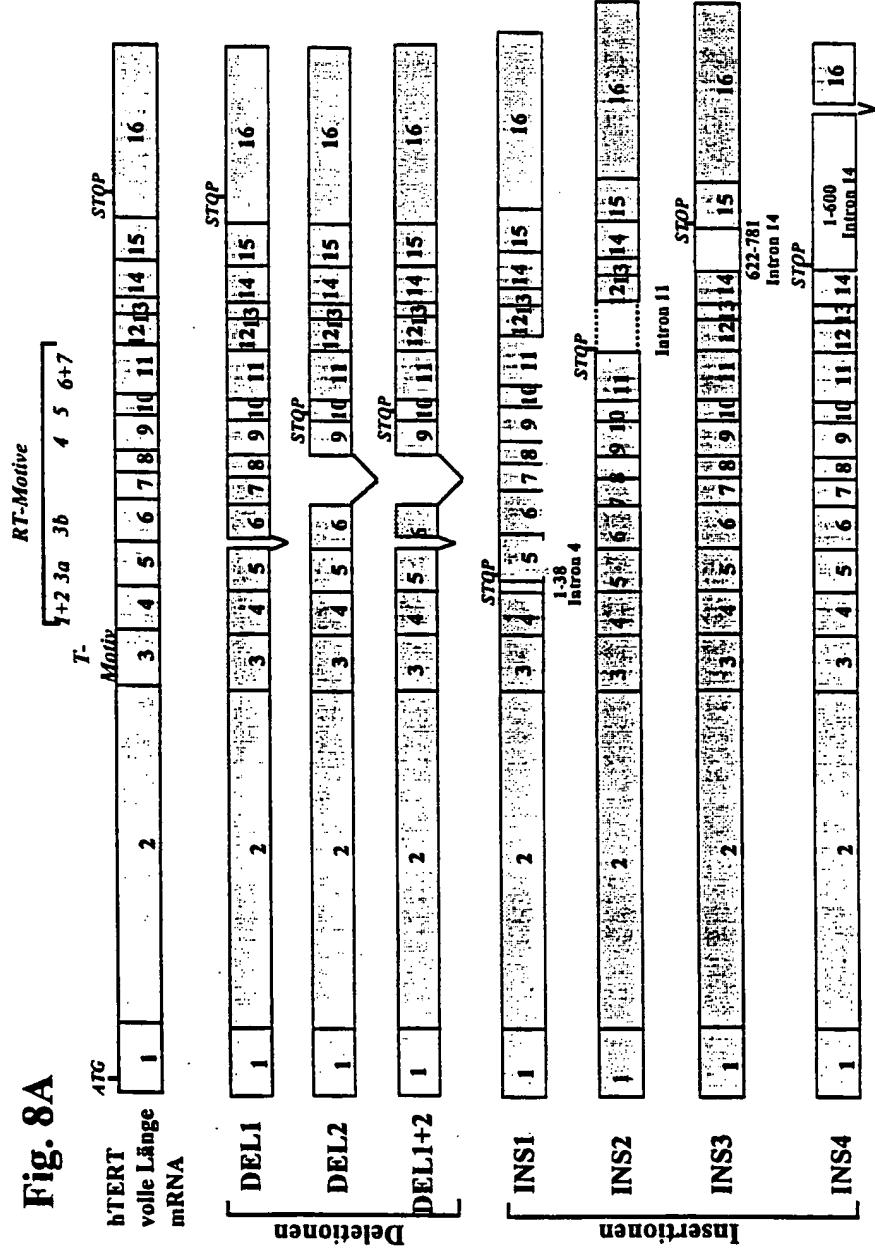
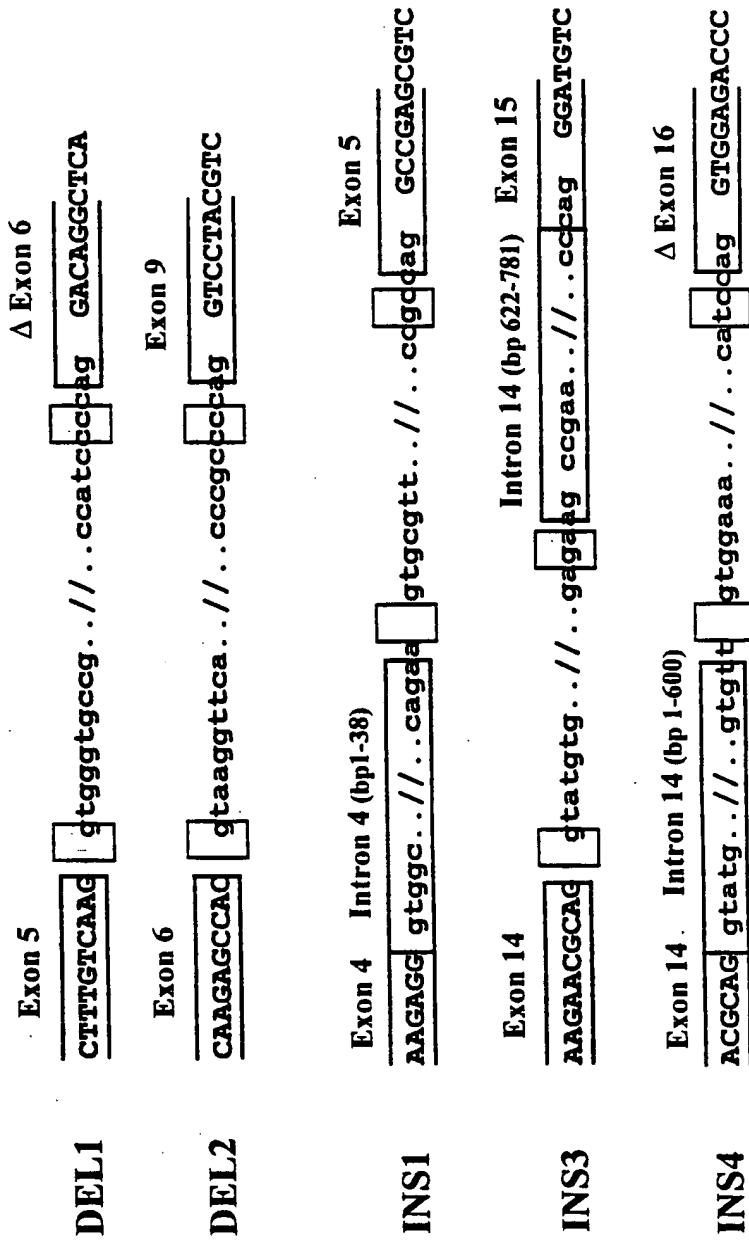


Fig. 8B



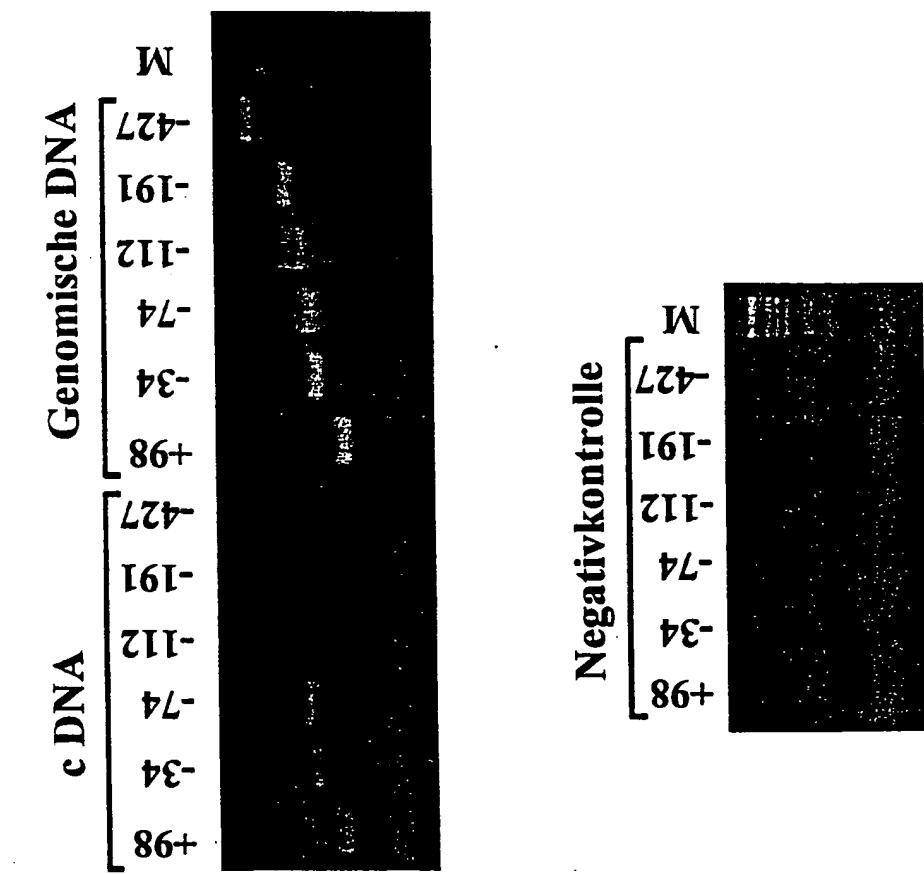


Fig. 9

Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCT GTCTAAAAAA AAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGG -11134
 ACAAAACAG AAATCAACAA CAAGAGGAAT TTGAAAACAT ATACAAACAC ATGAAAATTA AACAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAAAAGGA AATTGAAAAT TTATTAGA CAAATGATAA -10994
 CGGAAACATA ACCTCTCAAA ACCCACCGTA TACAGCAAA GCAGTGTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCCAGTGG CTCATGCCG TAACTCCAGC ACTTTGGAG -10854
 GCCAAGGCGG GCAGATGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTG -10784
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CTCGTAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGAT TGCCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTAAACCC CTGTCCTCAAG AAAAAAAA AAGTAGAAA ACTTAAAAT ACAACCTAAT -10574
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAA CCTAATGTTG GTAAAGAAA AGAAATAATA -10504
 AAGATCAGAG CAGAAAATTA TGAAACTGAA AGATAACAAT ACAAAAGATC ACAAAATTA AAAGTTGGTT -10434
 TTTGAAAAG ATAAACAAAAA TTGACAAACCC TTGCCCAGA CTAAGAAA AGGAAGAAG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAAG AGAGACATTA CAACTGATCA CACAGAAAATT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTCACATA ATAATTGAA AAACCTAGAA AAAATAGATA AATTCTCTAGA TGCATACAAC -10224
 CTACCAAGAT TGAACCATGA AGAAAATCCAA AGCCCAAAACCA GACCAATAAC AATAATGGGA TTAAAGGCCAT -10154
 AATAAAAAGT CTCTAGCAA AGAGAAGGCC AGGACCAAT GGCTTCCCTG CTGGATTTA CCAATCATTT -10084
 AAAGAAGAAT GAATTCCAAT CCTACTCAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACATCAT -10014
 TCTACATGGC CAGTATTAC CTTGATTCCAA AACCAGACAA AACACATCA AAAACAAACA AACAAAAAAA -9944
 CAGAAAGAAA GAAAATACAA GGCCAAATATC CTCGATGAAAT ACTGATACAA AAATCTCTAA CAAACACATA -9874
 GCAAAACAAA TTAAACAAACA CCTTCGAAAG ATCATTCTTGT GTGATCAAGT GGGATTATT CCAGGGATGG -9804
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCAAAC AAAATGAAGT ACAAAACATA -9734
 TATGATTATT TCACATTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCATGATA AAAACCTCA -9664
 AAAAACCCAGG TATACAGAA ACATACAGGC CAGGACACAGG GGCTCACACC TGCGATCCCA GCACTCTGG -9594
 AGGCCAAGGT GGGATGATT CTTGGGCCCC CTCAAAAAA CTTTTTTAA AAATTAGCCA GGCATGATGG CATATGCCCTG TAGTCCAGC TAGTCTGGAG -9454
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGACATG TCACTGTACT -9384
 CCAGCCTAGA CAACAGAACAGAAGCCACT GAATAAGAAG AAGGAGAAG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGG GGAGAAGTGG AGGGGGAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAAACATA TTCAACATA ATAAGGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGTCAA -9174
 AGCCTTCTCT CTAAGATCTG GAAAATGACA AGGGCCACT TTCACACTG TGATTCAACA TAGTACTAGA -9104
 AGTCTTAGCT AGAGCAATCA GATAAGAGAA AGAAAATAAA GGCATCCTAA CTGAAAGGA AGAAGTCAA -9034
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAA GACTTAAGAC ACCACTAAA AACTATTAGA -8964
 GCTGAAATTG GTTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTCTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAAGAAAC CAAAAGAGC GCTACAAATA AAATTAAACA GCTAGGATT AACCAAAGAA -8824
 GTGAAAGATC TCTACAAATGA AAATATAAA ATGTTGATAA AGAAATTTG AGAGGGCACA AAAAAGAAA -8754
 AGATATTCCA TGTTCATAGA TTGAGAAT ATAAACTGTG AAAATGTCCA TACTACCAA AGCAATTAC -8684
 AAATTCAATG CAATCCCTAT TAAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAACAA ATTCTAAGAT -8614
 TTGTCAGCAA CCACAAAGA CCCCAGAATAG CCAAAGCTAT CTCGACCAAA AAGAACAAA CTGGAAAGCAT -8544
 CACATTACCT GACTTCACAT TATACCTAAC AGCTATAGTA ACCCCAAACTA CATGGTACTG GCATAAAAAC -8474
 AGATGAGACA TGGACCCAGAG GAACAGAAATA GAGAATCCAG AACAAATTC ATGCATCTAC AGTGAACATCA -8404
 TTTTGACAA AGGTGCCAAG AACATACATTG GGGGAAAGA TAATCTCTT AATAATGGT GCTGGAGGA -8334
 CTGGATATCC ATATGCAAA TAACAATACT AGAAACTCTGT CTCTCACCAT ATACAAAAGC AAATCAAAT -8264
 GGATGAAAGG CTTAAATCTA AAACCTCAAAT CTTTGCACAT ACTAAAAGAA AACACCCGGAG AAACCTCTCA -8194
 GGACATGGGA GTGGCAGAAG ACTTCTTGAG TAATCTCTG CAGGCACAGG CAACCAAAGC AAAAACAGAC -8124
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCCTGACCA AGGAAACAAAT CAACAAAGAG AAGAGAACAC -8054
 CCACAGAATG GGAGAATATA TTGCAAAACT ATTCACTTAA CAAGGAATTA ATAACCAAGTA TATATAAGGA -7984
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAAGA TCTGGTAGA -7914
 CATTCTCAA AATAAGTCAT ACAAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAACAT ACTATGAGAC ATCATCTCAT CCCAGTAAA ATGGCTTTA TTCAAAGAC -7774
 AGGCAATAAC AATGCCAGT GAGGATGTGG TAAAAGGAA ACCCTTGAC ACTGTTGGT GGAATGGAAA -7704
 TTGCTACAC TATGGAGAAC AGTTGAAAG TTCTCTAAA AACTAAAAT AAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACTC CAAAAAAGGG AATCAGTGT TCAACAAAGCT ATCTCCACTC CCACATTAC -7564
 TGCAGCACTG TTCTACAGCAC CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAG -7494
 AAAATGTGGT GCACATCACAT ATGGGATGACT TACGCGGCCA TAAAAAGAA TGAGATCTG TCAGTTGCAA -7424
 CAGCATGGGG GGCACATGGTC AGTATGTTAA TGAAATAAG CCAGGCACAG AAAGACAAAC TTTCTATGTT -7354
 CTCCCTTACT TGTGGGAGCA AAAATTTAAA CAATTGACAT AGAAATAGAG GAGAATGGT GTTCTAGAGG -7284
 GGTGGGGAC AGGGTACTA GAGTCACACAA TAATTATTTG TATGTTAA AATAACTAAA AGAGTATAAT -7214
 TGGGTTGTTT GTAACACAAA GAAAGGATTAAT ATGCTGAAAG GTGACAGATA CCCCATTTAC CCTGATGTGA -7144
 TTATTACACA TTGCTATGCCCT GTATCAAAT ATCTCATGTA TGCTATAGAT ATAAACCTA CTATATTAAA -7074
 AATTAAAATT TTAATGGCCA GGCACCGTGG CTCTATGCTT CAATCCAGC ACTTTGGAG GCGGAGGCGG -7004
 GTGGATCACC TGAGGTCAAG AGTTGAAAC CAGTCTGGCC ACCATGATGA AACCCCTGTCT CTACTAAAGA -6934
 TACAAAAATT AGCCAGGGCGT GGTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTTGAAC CTGGGAGGCG GAGGTTGCAAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAAAACAA AAAAAGAAAG ATTTAAATTG TAATTCTTAT GTACCGTATA -6724
 AATATATAC TCTATATATT AGAAGTAAA AATTAACAA ATTAAACAG GTAAATACCA ACTTATCTA -6654
 AAATAAGAAC AATGTATGTG GGGTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTAAACC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAA ATCGAGCTGC AGAATTGGCA CGCTCTGATCA -6374
 CACCGCTCTC TCATTCACGG TGCTTCTTT CTGAGGATTT CGATTGTTG TTGCTGTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTTGACCTG AGCCACTCA ATCTTCAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCGAG AGGGCTGTG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCTGCCT -6094
 CGTGTGGTGGAG CAGCGATGAG AGTGCCTTAA TTTACGCTTT GCAAAGATTG CTCTGATAC CATCTGGAAA -6024
 AGGGCGGCCAG CGGGATGCA AGGAGTCAGA AGCCTCTGC TCAAACCCAG GCCAGCAGT ATGGCGCCCA -5954
 CCCGGGGCTG TGCCAGAGGG AGGAGGTCAGA AGGCACCTCG AAGTATGCT TAAATTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGATT CTGAGGGAAAG CTTGAGTTAG GTGCTCTTT TAAACAGAA ACTCATGAA -5814
 GCACCCCTCT CAAGGGAAAA CCAGACGCC GCTCTGCGGT CATTACCTC TTTCCTCTCT CCCTCTCTT -5744
 CCCCTGGCTT TTCTGATCGG GACAGAGTGA CCCCCGTGGA GCTTCTCCGA GCCCCGTGCTG AGGACCCCTCT -5674
 TGCAAAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCTCTG CTTAAATACA AACTGGGATG -5604
 TGGCTGGGG CGGACAGCGA CGGGGGGATT CAAAGACTTA ATTCCATGAG TAAATTCACAT CTTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAATTT TCTAAATTAAC ACATTCAGGA CTGCAAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTCGCC CTAAGTACTT TTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGGCTGG CGGGCAGGGC TATGAGCAGC AGCAGGGCCAG CGGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CTCCTCTGG AGTGCACCA CAAGGGCTGGC CACGCTGCCT -5184
 GTGACTCAGG ACCCCATACC GGCTTCTGG GCCCACCCAC ACTAACCCAG GAAGTCAAGG AGCTCTGAAC -5114
 CCGTGGAAAC GAACATGACC TTGCTGTGCG TGCTTCCCTG GGTGGTCAA GGGTAATGAA GTGGTGTGCA -5044
 GGAATGGCC ATGAAATTAA CACGACTCTG CTGATGGGGA CCGTCTCTTC CATCATTATT CATCTTCACC -4974
 CCCAAGGACT GAATGATTCTC AGCAACTTCTC TCGGGTGTGA CAAGCCATGA CAAAACCTAG TACAACACACC -4904
 ACTCTTTAC TAGGCCCCACA GAGCACGGSC CACACCCCTG ATATATAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTCTAGGC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTCTC TCTAGAGTAG TAGACCCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCTGGT TGCTGCTTCC CGAGGGGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGTGCG ACAGTGGAGC CAGCCCTGTC TCCACACCTC CCGCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
 TTCCCTAACCTC CTGGGTGGGC CGTGTCTCCAG CGCTACTGTC TCACCTGTCC CACTGTGCTC TGCTCAGCG -4554
 ACGTACTCG CACGGCTCTC CCTCACATGG GGTGCTGTG TCCTTCCCA ACACCATGAC GCGTTGAAGG -4484
 GAGGAGATTG TGCCCTCTCC AGACTGGCTC CTCTGAGCTT GAAACCTGGG CAGGGGGCCC GATGCAAGGTT -4414
 CCTGGCTCC GGCTGCACGC TGACCTCCAT TTCCAGGCC TCCCCGTCTC CTGTCATCTG CCGGGGCTG -4344
 CCGGTGTGTT CTCTGTGTT TGTCCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCG CTAGGGTCTC -4274
 GGGGTTTTA TAGGCATGG AGGGGGGGT GTGAGGATTTA GGTGGGCAAG GGCCTCTTG GGAAATGCAA CATTGGGTG -4204
 TGAAGTAGG AGTGCCTGCT CTACCTCTAGG TCCACGGGCA CAGGCCCTGG GATGGAGCCC CCGCCAGGAA -4134
 CCCGCCCTTC TCTCCCAGC ACTTTCTCTG CCCCCCTCCCT CTGGAAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AGAACCCAGC ATTGGCACCC CTGGACATT GCCCCACAGC CCTGGGAAATT -3994

c-Myc

CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTATT TAAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACAAAC TGGTTAAACAA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCAAGTGAAG GAGGAACATG CCGTTATAAG AGCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAC TGCCACCTCC ATGGGATACG TACCCAACAT GCTCAAAAG AAAGAATTTC ACCCCATGGC -3714
 AGGGGAGTGG TTAGGGGGT TAAGGACGGT GGGGGCGCA GCTGGGGCT ACTGCACGCA CCTTTTACTA -3644
 AAGGCACTT CCTGGTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGAG TGGGGATGGG -3574
 GGAACCCCGGA GGCTGTGCCA TCTTGTCCAT GCCCCAGTGT CCTGGGAGGG ATAATGCTCT AGAGATGCC -3504
 ACGTCTCTGAT TCCCCAAAC CTGTTGACAG AACCAGGCC CTTTGAGGT GTGATCTCCG -3434
 TGAGGACCTT GAGGTCTGGG ATCTCTCTGG ACTACCTGCA GGCCCCAAAAGA GTAACTCAGG GGTGCTGGGA -3364
 AGAGGGCGGC AGGAGGGTCA GAGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGGCC AGGGCTGCAA GCGCTTCCAG AAGCTGAAA AAGCAGGGAA GGGACCCCTCC -3224
 ACGGAGCTG CAGCAGGAAG GCACGGCTGG CCCTTAGGCC ACCAGGGCC CTCGGTGGACC TCCGGCTCTC -3154
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCACT GTGAACTCTA GGATTATTC AAAACAAAGG TTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
 GTGCCCTCCGG GCAAGGGCAG GGCAGGCAG AGTGTATT TTTAGCTATT TTATTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTTGCCAG GCTGGAGTGC AGCGGCATGA TCTTGCTCA CTGCAACCTC -2874
 CGTCTCTGG GTTCAGCAA TTCTCTGTC TCAGCTCTC AAGTAGCTGG GATTTGAGGC GTGACCCACC -2804
 ACACCCGGGT AATTGGTAT TTGAGTGAAG GATGGCTT CACCATGTT GTCAAGCTGA TCTCAAATC -2734
 CTGACCTCAG GTGATGCCAG CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCTTATTTA ACCATTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTCCCTT TAATCTGAGA GTTACCTCTC TTGATATT TCTGTAATT TCTGTAATT CTCGTAAGACT GGGGATACAC -2524
 CGTCTCTTGA CATATTCACA CCAGCTGTGA CCAGCTGTGA TCCCATGGGA CCCACTGCAAG GGGCAGCTGG -2454
 GAGGCTGCAAG GCTTCAGGTC CCAGTGGGGT TGCCCATGTC CAGTAGAAAC CTGATGTTAGA ATCAGGGCC -2384
 AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTGAGTGTG GCTGAACAT GTAGAAATTAA AAGTCCATCC -2314
 CTCTACTCT ACTGGGATTG AGGCCCTTCC CTATCCCCC CCAGGGCAG AGGAGTCTC CTCACTCTG -2244
 TGGAGGAAGG AATGATCTCT TTGTTATTCTT CACTGCTGT ACTGAATCCA CTGTTTCATT TGTGGTTG -2174
 TTGTTCTGTTG TTGAGAGGG GGTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGAATGG CGCGATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTCA AGTGAATCTC CTGCTCCCG CTCCCAATTG GCTGGGATTA -2034
 CAGGCAACCCG CCACCATGCC CAGCTAATTG TTGTTATTG TAGTAGAGAC GGGGGTGGGT GGGGTTCAAC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
 GATTACAGGT GTGAGGCCACC ATGCCAGCT CAGAATTAC TCTGTTAGA AACATCTGGG TCTGAGGTAG -1824
~~CAAT-Box~~
 GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTATTGT TGTTAGAAC -1754
 CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTCAAAG ACACACTAAC TGCAACCCATA -1684
 ATACTGGGGT GTCTCTGGG TATCAGCAAT CTTCATGAA TGCCGGGAGG CGTTCCCTCG CCATGCACAT -1614
 GGTGTTAATT ACTCCAGCAT AATCTCTGC TTCCATTCTCT TCTCTCCCT CTTTAAAAT TGTGTTTCT -1544
 ATGTTGGCTT CTCTGCAGAG ACCAGTGTAG AGCTACAAC TAACTTTGT TGGAACAAAT TTTCAAAC ~~CC~~ -1474
~~Spi~~
~~GCC~~ CTTGC CCTAGTGGCA GAGACAATTCAACACAG CCCTTTAAAAA AGGCTTAGGG ATCACTAAGG -1404
 GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
 GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTCCTCC GGCAGTTCT -1264
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTCAGTGT TTGCCGACCT CAGCTACAGC -1194
 ATCCCTGCAA GGCTCGGGGA GACCCAGAAG TTTCTCGCCCT CTTTAGATCC AAACTTGAGC AACCCGGAGT -1124
 CTGGATTCCCT GGGAAAGTCCT CAGCTGTCCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCAAGTGG -1054
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCCAGTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
 GCCTGGACCC CGAGGGCTGCC CTCCACCCCTG TGCGGGCGGG ATGTGACCAAG ATGTTGGCCT CATCTGCCAG -914
 ACAGAGTGCC GGGGCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCCG CCGGTGGCGC GCCAGCAGGA -844
~~CCAC-Box~~
~~Spi~~
 GCGCTTGGCT CCATT~~TT~~CCCA CCGTTCTCG ACAGGGACCGC CCGGTGGGT GATTAACAGA TTTGGGGTGG -774
 TTTGCTCATG GTGGGGACCC CTCGGCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGGT GCGCGTCCAG GGAGCAATGC -634
~~AP-2~~
 GTCTCTGGGT TCGTCCCCAG CCGGTCTAC GCGCCTCCCGT CCTCCCTTC ACGTCCGGCA TTCGTGGTGC -564
 CCGGAGCCCG ACGGCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
 GGTCCCGCAG CGCACCTGTT CCCAGGGCCT CCACATCATG GCGCCCTCCCT CGGGTTACCC CACAGCCTAG -424
~~Spi~~
 GCGGATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCC GGGAGGCCA GCGG~~CGCG~~ -354
~~Spi~~
~~GGCGGGAAAG CGCGGCCAG ACCCCCCGGT CCGCCGGAG CAGCTGGCCT GTCGGGGCCA GGCGGGCTC~~ -284
~~c-Myc~~
 CCAGTGGATT CGGGGCACA GACGCCAGG ACCGCGCTCC ~~CCACGTGGCG~~ GAGGGACTGG GGACCCGGGC -214
~~Spi~~
 ACCCGTCTG CCCCTTCACC TTCCAGCTCC GCGCCCTCCG CGGGACCC ~~GCCCCCTCCC~~ GACCCCTCCC -144
~~Spi~~
 GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCC ~~CCCCCTCC~~ TCCGCGCC ~~CCCCCTCC~~ -74
~~TCGCGCGCG AGTTTCAGGC AGCGCTGCCT CCTGCTGC~~ ~~ACGTGGAAAG CCCTGGCCCC GGCCACCCCC~~ -4
 GCGATG

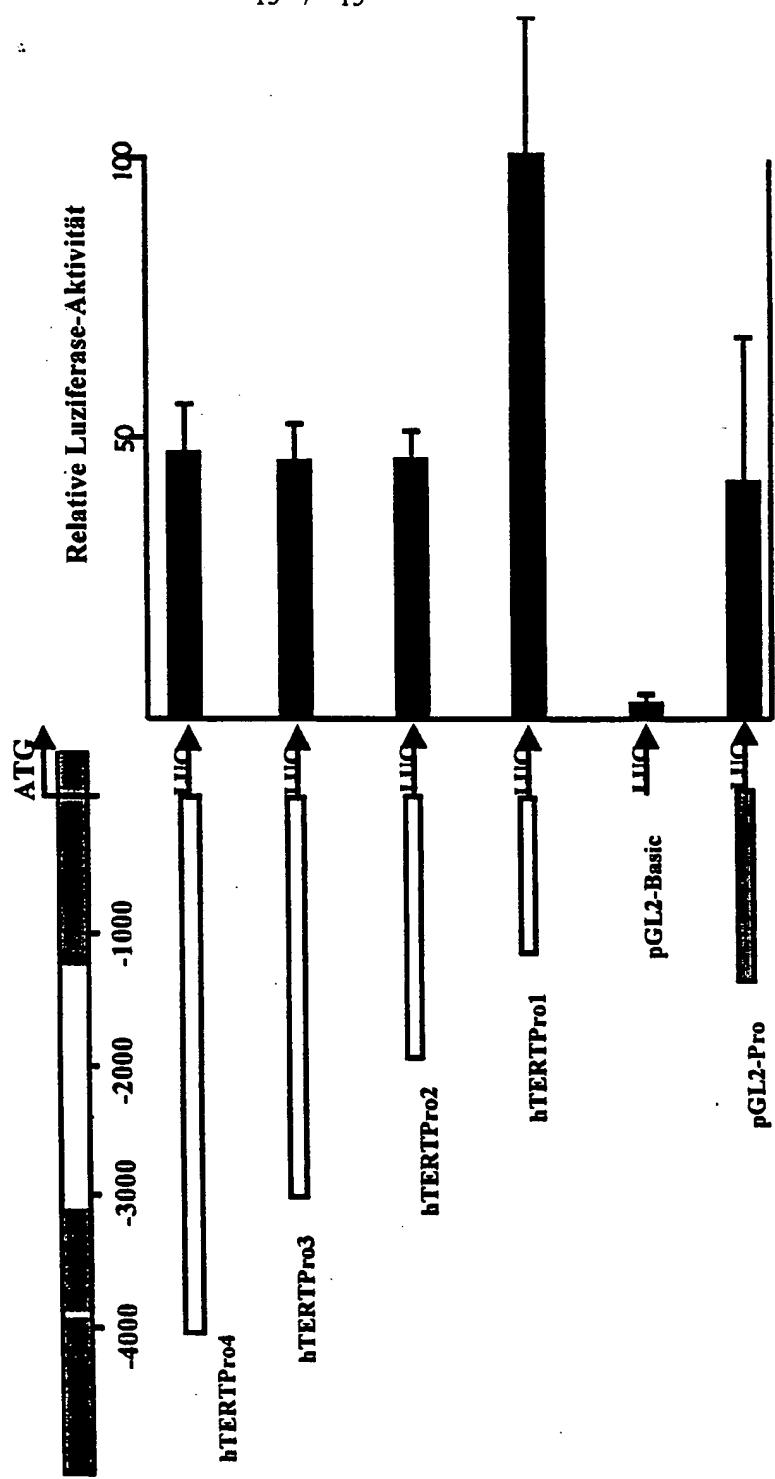


Fig.: 11

SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

15 <140>
 <141>

20 <160> 20

25 <170> PatentIn Vers. 2.0

30 <210> 1
 <211> 5126
 <212> DNA
 <213> Homo sapiens

35 <400> 1
 gagctctgaa ccgtggaaac gaacatgacc cttgectgcc tgctccctg ggtgggtcaa 60
 gggtaatgaa gttgggtgtca ggaaatggcc atgtaaattt caccgactctg ctgatgggg 120
 ccgttcccttc catcattatt catcttcacc cccaaaggact gaatgattcc agcaacttct 180
 tcgggtgtga caagccatga caaaacttcag tacaaacacc actctttac tagggccaca 240
 gagcacgggc cacacccctg atatattaag agtccaggag agatgaggct gctttcagcc 300
 accaggctgg ggtgacaaca gcggcgtgaac agtctgttcc tcttagactag tagaccctgg 360
 caggactcc cccaaattctt agggccctgtt tgctgttcc cgaggggcgcc atctgcctgg 420
 gagactcagc ctgggggtgc acactgaggc cagccctgtc tccacacccct cccgcctccag 480
 gcctcagtt ctccagcagc ttccctaaacc ctgggtggc cgtgttccag cgctactgtc 540
 tccacctgtcc cactgtgtct tgcgttcagcg acgtgttcgtc caccgttctt ctcacatgg 600
 ggtgtctgtc tccctcccca acacttcacat ggggttggaaagg gaggagattc tgccgcctccc 660
 agactgggtc tcttgagccctt gaaacctggct cgtggccccc gatggcgggtt cctggcgttcc 720
 ggctgcacgc tgacctccat tttccaggcgc tccccgttcc ctgtcatctg cccggggctgg 780
 ccgggtgttt tttctgtttt tccacgttca gtcgcgtgtc tctctgcctcg 840
 ctgggggttc ggggttttta taggcatagg acggggggcgt ggtggccag ggcgcgttcc 900
 ggaacatgaa catttgggtt gggaaatggg tagtgcgttc ctcacatgg tccacgggca 960
 caggcctggg gatggagccc ccggcaggag cccgccttcc tctgcggcagc actttcttgc 1020
 cccgccttccct ctgaaacaca gatggcgtt tttccacaacg actaagcatac ctcttcacaa 1080
 aagacccagc attggcaccc ctggacattt gccccacagc cctggaaatt cacgtgacta 1140
 cgcacatcat gtacacactt ccgtccacga ccgacccccc ctgttttttt ttaatagcta 1200
 caaaacggagg aaatccctgc taaaatgtcc ttaaacaac tggttaaaca aacgggttcca 1260
 tccgcacggc ggacatgttcc tcaacatgtt gggaaatggc cctgttataaa agcctgcagg 1320
 catctcaagg gaaatgttca gatgtttttt tccacatgttcc atgggatatacg tacgcaacat 1380
 gctcaaaaag aaagaatttt accccatggc aggggatgg ttaggggggt taaggacgg 1440
 gggggcggca gctgggggtt actgcacgca ctttttacta aagccatgtt cctgggttcc 1500
 atggatattgg ctcaatgtt gggactaaac catagggggg tggggatggg ggaacccgg 1560
 ggctgttccca tctttggccat gcccggatgt cttggcagg ataatgttctt agatgtcccc 1620
 acgtccctgtt tcccccaaaac ctgtggacag aacccggggc gccccggggc ctttgcagg 1680
 gtgtatctccg tgaggacccctt gagggtctggg atctttccggg actacatgtca gccccggaaa 1740
 gtaatccagg ggttctggg agaggcgggc aggagggtca gaggggggca gcctcaggac 1800
 gatggaggca gtcagtctg ggctgaaaag ggaggggggg cttcgagccc aggctgcata 1860
 ggcgcctccca aagctggaaa aageggggaa gggacccctcc acggagctg cagcaggaag 1920
 gcacggctgg cccttagccc accaggggcc atcgtggacc tccggccccc gtgcctatgg 1980
 agggcactcg cgctggccctt ctacatgtaa gtgtgtggg atttgcagaa gcaacaggaa 2040
 acccatgtcac tttgtatcttta ggattatttc aaaacaaagg tttacagaaa catccaaggaa 2100
 cagggtctgaa gtgcctccgg gcaaggggcag ggcaggcacc agtgattttta tttagcttt 2160
 ttatatttttacttctt ctgagacaga gttatgttctt tggtggccag gctggagttc 2220
 agcggcatgtt ctgggttccca ctgcaacccctc cgtctcttgg gttcaagcaa ttccgttcc 2280
 tcagctccca aagtagctgg gatttccaggc gtgcacccacc acacccggctt aattttgtat 2340
 ttttagtata gatgggtttt caccatgtt gtcagatgtca tctcaaaaatc ctgacccatgt 2400
 gtgtatccggc cacccatgttcccaaaatgtt ctgggattac aggcatgtac cactgcaccc 2460

atacaaaacac atgaaaatata aacaatatac ttctgaatga ccagttagtc aatgaagaaa 240
 ttaaaaagga aattaaaaaa ttatattaag caatgataa cgaaacata accttc当地 300
 acccaggtt tacagcaaa gcagtgttaa gaaggaagtt tatagtata agcagctaca 360
 tcaaaaaagt agaaaagcca ggcgcagtgg ctcatgctg taatccc当地 accttgggag 420
 gccaaggcgg gcagatccccc tgaggctcagg agttc当地 cagcctgacc aacacagaga 480
 aacccctgtc ctactaaaaa tacaaaataa gctgggcatg gtggcacatg cctgtatcc 540
 cagctactcg ggaggctgag gcaggataac cgctt当地 caggaggctgg aggttgcgg 600
 gagccgggat tgcccatg gactccggcc tggtaacaa gatgtt当地 ctgtt当地 660
 aaaaaaaaaa aagttagaaaa acttaaaaat acaacccata gatgcacccctt aaagaacttag 720
 aaaaggcaaga gcaaaactaaa cctaaaaattt gtaaaaagaaa agaaaataata aagatcagag 780
 cagaaaaataa taaaaactgaa agataacaaat acaaaaatggc aacaaaatata aaagtggg 840
 ttttggaaag ataaaacaaa ttgacaaacc tttgccc当地 ctaagaaaaa agggaaaag 900
 acctaaataa ataaaatgtc agatgaaaag agagacattt caactgatac cacagaaaatt 960
 caaaggatca cttagggcata ctatgagca ctgtacacta ataaaatggaa aaacccctgaa 1020
 aaaatagata aattctt当地 tgcatacatac ctaccaatg tgaaccatgaa agaaaatccaa 1080
 agcccaaaa gccaataac aataatgggaa ttaaaggccat aataaaaaatg ctccctgaa 1140
 agagaaggccc aggacccatg ggctt当地 ctggatctt当地 ccaatcattt aaagaagaat 1200
 gaattccat cctactcaaa ctatctgaa aataagagga aagataactt ccaaaactcat 1260
 tctacatggc ctagtattacc ctgattccaa aaccagacaa aaacacatca aaaaacaaaaca 1320
 aacaaaaaaa cagaagaaaa gaaaactaca ggccaaatatac cctgtatgaa actgatacaa 1380
 aaatctt当地 caaaacacta gcaaaacccaa ttaaaccataa cctt当地 cgg 1440
 gtgtatcaatg gggatattt ccaggatggg aaggatgggaa caacatatgc aaatcaatca 1500
 atgtgatatac tcatcccaac aaaaatgttata ttttgggatggg 1560
 cagaaaaaaggc atttggataaa attctgc当地 cttcatgata aaaaacccctt aaaaaccagg 1620
 tatacaagaa acatacaggc caggcacatg ggctc当地 ctc当地 gcaacaaaat 1680
 aggccaaagggt gggatgggaa ctttggccca ggagttt当地 actagctgg gcaacaaaat 1740
 gagacccctggt ctacaaaaaa cttttt当地 aaatttggcc ggc当地 gatggg 1800
 tagtccccag tagtctggag gctggatggg gagaatcact taaggctt当地 aggttggag 1860
 tgc当地 gggc当地 tcaatctt当地 ccaggctt当地 caacacaaa agacccctt 1920
 gaataagaag aaggagaaggg agaaggggaga agggaggggag aaggaggagg gaggagaagg 1980
 aggagggtggg ggagaagggttgg aaggggaaagg ggaaggggaaa gaggaagaag aagaaacata 2040
 tttcaacata ataaaaggccc tatatgacag accggatggtag tattatgagg aaaaactgaa 2100
 agccctt当地 ctaatgtc gaaaatgaca agggccactt ttc当地 actt当地 tgatcaaca 2160
 tagtactaga agtccctatg agaactt当地 gataagagaa aaaaatggggaa ggc当地 2220
 ctggaaaggc agaaggctt当地 ttatctt当地 tgcatgatgat atgtt当地 atctggggaaa 2280
 gacttaagac accactaaaaa aactt当地 gctgaaatggt ggtacagc当地 gataaaaaat 2340
 caatgtacaa aaatctt当地 tttt当地 ttccaaacagc aaacaatctg aaaaagaaaac 2400
 caaaaaaggc gctacaaaata aatgggatgg aaccaaggaa gtggaaatggc 2460
 tctacatggc aaaaatggataa atgtt当地 aaaaatggc aagggccaca aaaaagaaaat 2520
 agatatttca tggatcatgaa ttggaaatgg aatactt当地 aaaaatgtcc tactacccaa 2580
 agcaatttac aatctt当地 caatctt当地 taaaatacta atgactt当地 tcacagaaaat 2640
 agaagaaaaca attctt当地 atgtt当地 ggatggatggcc caacaaaatggc ccaaaactat 2700
 cctgacccaa aagaacaaaatggc acatggatgg gactt当地 tataactacaa 2760
 agctatagta accccaaacta catgtt当地 gctatgggaaatggc agatggatgg tggacccagg 2820
 gaacacataa gagaatggc aaacaaaatggc atgc当地 atgtt当地 tttt当地 2880
 aggtgccaatg aacatactt当地 gggggaaaatggc taatctt当地 aataatgtt gctgggggg 2940
 ctggatatcc atatgcaaa taacaatact agaactt当地 ctctc当地 atacaaaagc 3000
 aatctt当地 ggatggaaaatggc ttggatggc aacatctt当地 ctggatggactt actaaaagaa 3060
 aacccggggaaatggc aacatctt当地 ggacatggc gtggggaaatggc acctt当地 ctggatggactt 3120
 caggccatggc caacaaaatggc aatggggatggc atataatgtt aaaaatgtcc 3180
 tggccatggc agggaaaatggc caacaaaatggc aacatctt当地 ctggatggactt 3240
 ttggatggactt atctcatctt当地 caaggatggc atacaaaatggc gctt当地 3300
 ctctataaga aaaaacccatggc ataaatgtt当地 aacatctt当地 tttt当地 3360
 catttctt当地 aataatgtt当地 aacatctt当地 tttt当地 3420
 ctgtatcatgaa gagaatggc aatctt当地 actatgtt当地 atctcatctt当地 3480
 atggctt当地 ttggatggactt aacatctt当地 atgtt当地 gggatggactt 3540
 accctt当地 gggatggactt aacatctt当地 tttt当地 3600
 ttggatggactt aacatctt当地 tttt当地 3660
 aactaaaaatggc aatctt当地 tttt当地 3720
 ccaacatctt当地 tttt当地 3780
 aaaaatgtt当地 gcaatctt当地 aatggggatggc aacatctt当地 tttt当地 3840
 tttt当地 tttt当地 3900
 aacatctt当地 tttt当地 3960
 aacatctt当地 tttt当地 4020
 aatctt当地 tttt当地 4080

5	ctcttagtaat	ctagtaattc	tttttttaaa	ttgtctttag	tactgccaca	ctggggcttc	3600
	tttgattagt	atttctgc	tgtgtctgtt	ttctgcctt	aatttatata	tatatatata	3660
	ttttttttt	tttgagaca	gagtcttgg	ctgtcgccca	gggtgagtgc	agtgggtgtga	3720
	tcacaggctca	gtgtaaactt	taccccttgc	cctgaggccgt	ccttcacccat	cagcccttcgt	3780
	agtagcttga	actgcagaca	ccgcacccgt	cacccgtta	attttaaat	tttttttgg	3840
	gacagggtct	tgctgttgg	ccggccgtgg	tctcaactc	ttggactcaa	gggatccatc	3900
10	tacctcggtct	tcccaaaatgt	ctgaattaca	ggcatgagcc	accatgtcg	gcctatattt	3960
	caacactttt	atattttat	agtgtggta	tgtcctgtta	acagcatgt	ggtaatttc	4020
	caatccagtc	tgacagtcgt	tggttaactg	gataacctga	tttattttca	tttttttgtc	4080
	actagagacc	cgccctggcgc	actctgttgc	tccactgtc	tgtgtcatgt	cctctgttccc	4140
	ttgttctctca	ccacacttgc	gggttgcctg	tgegttctt	ccggactgtg	tgttgatect	4200
	ctcggtgtct	cctgggtact	gggcattttgc	ttttttttt	tttttttttgc	tttttttttgc	4260
	tgatctttt	atgtcggtt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4320
15	ggctggagtg	taatggcaca	atctcggttc	actgcaccc	ctgccttc	ggttcaagca	4380
	gttctcatc	ctcaacctca	tgagtagctg	ggattacagg	cgcccaccac	cacgcctggc	4440
	taatttttgt	attttttagt	gagataggt	ttcaccatgt	ttggccagggt	ggttcaaac	4500
	tcttgaccc	aagtgtatcg	ccgccttgg	ccctcccaag	tgtctggatt	acagggtcaa	4560
	gecaccgtgc	ccggcatacc	ttgtatctt	aaaaatgt	ctgaaacatt	gttaccccttgc	4620
20	tcttgagcaa	taagaccctt	agtgatttt	agttttttttt	tttttttttgc	tttttttttgc	4680
	ttttccctgc	tgacttagt	ctatctcagg	catcttgc	ccccccacaag	ctaaagcattt	4740
	ttaatattgt	tttccgttgc	gagtgtttt	gtagtttgc	ccccccccc	tttttttttgc	4800
	tttgttcccc	gttgttcttc	tggttcttc	ccggccgttgc	gggttccccc	tttttttttgc	4860
	tgcgtgttgc	ttctgttctt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4920
25	atggcattca	gacacgttcc	ggggaccccttgc	tttttttttgc	tttttttttgc	tttttttttgc	4980
	tacacgaggag	ggcggtcattc	tttggcccttgc	agtgcttgg	gcacccacgt	ggccacgttc	5040
	cttagccagt	gagtgtacgc	aacgtccgt	ccggcccttgc	tcagcttgc	aaaccccccagg	5100
	catgtcgggg	tctgggttgt	ccgggggtgtc	gagtttgc	tcgcgcacac	ctgggtgtgc	5160
	gecggccatgc	tgacgggtgt	gcttgggggg	ggagtgttgc	tttttttttgc	tttttttttgc	5220
30	aaccaggaca	aaggatgagg	cttgcggcc	tttttttttgc	tttttttttgc	tttttttttgc	5280
	atgtggataa	tttttttttgc	tctgggttgc	ggcgggttgc	tcacgccttgc	aatttttttgc	5340
	ctttgggagg	ccaaaggccgg	tggatcacgt	ggtcaggagg	tcgagaccat	cttggccaaac	5400
	atgtatggaa	cccatctgt	ctaaaaacac	aaaaatttgc	tttttttttgc	tttttttttgc	5460
	tgtatccca	gttactcggtt	aggctggaggc	aggaaatttgc	tttttttttgc	tttttttttgc	5520
35	gttgcagtga	ggcgacatttgc	caccacttgc	tttttttttgc	tttttttttgc	tttttttttgc	5580
	ctcaaaaaaaaa	aaaaaaaaaa	aaaaatttgc	tttttttttgc	tttttttttgc	tttttttttgc	5640
	aagaaaaagg	aaaaaaatgt	ttataatagg	tttttttttgc	tttttttttgc	tttttttttgc	5700
	atcattttag	gggttatttgc	gtggggagcat	cacttcacagg	acatttgc	tttttttttgc	5760
	tttgttctgc	ggatcccttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	5820
40	ggcttccatc	ggccatggc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	5880
	cccttcagtgt	gctggatgt	cagtgttcc	atgggtgcac	tttttttttgc	tttttttttgc	5940
	ccctgtgtgt	agttggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6000
	ggccctcggt	atggggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6060
	ggccctgtgt	gagctggat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6120
45	ggccctcggt	aagctggagg	tatggatgt	ggatgtgt	tttttttttgc	tttttttttgc	6180
	ggccctcggt	tgagctggat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6240
	agggccctcg	gtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6300
	cagacgggtgc	cagaccatgc	ggtgatgt	tttttttttgc	tttttttttgc	tttttttttgc	6360
	ggtgatgtgt	ccaggccctgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6420
50	gtgtgagggtc	accaggccct	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6480
	gggggtgaagg	tcgcccaggcc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6540
	gtctggatgt	aggttcggcc	ggccctcggt	tttttttttgc	tttttttttgc	tttttttttgc	6600
	gtccgggggt	aggttcggcc	acccttcgggt	tttttttttgc	tttttttttgc	tttttttttgc	6660
	gggttggatgt	gggttggatgt	ggccctcggt	tttttttttgc	tttttttttgc	tttttttttgc	6720
55	ggtccgggggt	gagggtccca	gaccctcggt	tttttttttgc	tttttttttgc	tttttttttgc	6780
	agggttggatgt	tgaggttcggcc	agaccctgt	tttttttttgc	tttttttttgc	tttttttttgc	6840
	cagggttcagg	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6900
	cagggttcagg	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	6960
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7020
60	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7080
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7140
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7200
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7260
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7320
65	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7380
	ttgtggatgt	ttgtggatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	7440


```

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens
5
<400> 7
gtggctgtc tttggtttaa cttcctttt aaacagaagt gcgttgagc cccacatgg 60
gtatcagctt agatqaagggg cccqgaggag gggccacggg acacagccag gccatggca 120
cgccgcacaa ccatgggtgc gcacagtggag gtggccgagg tccgggtgcc tccagaaaaag 180
10 cagcggtggg gtgttaggggg agctcctggg gcaggagacag gctctgagga ccacaagaag 240
cagccggcc agggcctgga tgcaccaacgg cccgaggtcc tggatccgtg tccctgtgt 300
gtgcgcagcc tccgtcgctt cccgttacg gggctggggg accaggccac gactgcccagg 360
agcccacccgg gctctgagga tccgtggaccc tggccacgg ctccgtcacc ccacccctgt 420
ggctcggtg gctcggtgaa ccccgatc tggaggaggt gtgggggttag gtggacagag 480
15 gtgtggatg aggatccccgt gtgcacacaca catgcggcca ggaacccgtt tcaaacaggg 540
tctgagggaaat ctggggagggg ttcttaggtcc cgggtctggg tggctgggaa cactggggag 600
gggctgttcc tccccctgggtt ccctatgggtt ggggtgggac tggccggat ccactttctt 660
gactgtctcc catgtgttcc cggcccaag 687
20
<210> 8
<211> 494
<212> DNA
<213> Homo sapiens
25
<400> 8
gtgggtgccc gggacccccc tgagcagccc tgctggaccc tgggagtggc tgctgtattg 60
gcacccatcg ttgggtggag gaggtactcc tgggtggggcc gcaggaggtg cagggtaccc 120
tgtcaactgtt gaggacacac ctggcaccta gggtgaggcc ctccacccctt tccctgcagca 180
30 catggggccctt actgtgcacc ctgactgttcc gggctcttat tcccaaggag ggtcccaactg 240
gatccccatgt tccgtcgatg aaggaaacccg aacgggtcag ccaccaggcc cccgtgcctt 300
gcaccccaatgtt cttggggccctt ggggtctctg tccctgaggct cagagagggg acacagcccg 360
ccctgccttcc ggggtctggg tgggtggggg ttagagagag agtgggggac accggccaggc 420
caggccctgaa gggcagaggtt gatgtctgag tttctgtgtt gccactgtca gtctccctcgc 480
494
35
<210> 9
<211> 865
<212> DNA
<213> Homo sapiens
40
<400> 9
gttaagggttca cgtgtgtatag tccgtgtccag gatgtgtgtc tctggatataatgtgtct 60
agaatgcagt cgtgtctgtt atgcgtttt gttggggagg tacttccatg atttacacat 120
45 ctgtgtatgt cgtgtgtggc acgtgtgtgtt cttgggtgtat gtatctgtgg cgtgtcatatt 180
tgtgtgtgtt gtgtgtgtgg cactgtgtgtc tccatgtgtt gttgggtgtt ggtgtgtcatg 240
tgtgtgtgtt gtgtgtgtcat gcatgtgttcat gctgtgtgtt gcatgtgttcat gatgtgtgtt 300
tttgtgtgtt gtgtgtgtcat gtgtccgttcatatgtgtgtc tctatggcat ggggtgtgtt 360
ggcccccgttgg ctttacttccat tccctccatca ggcattgtcc gcaccattgtt ccttacgttc 420
tcgggtgtctg gttttggggag ctccacatcc agggcccttca cttttagcat ggggtggccct 480
50 gtcctgttcaac agggctgggc ctggagact gtaagccagg tttgagagga gatgtggat 540
gctgggtgttcaac ctttccgttca gcccgttgcac ccccaaggacc ctatggccctt 600
tccatgtatggat ataggaaaggc tgatgttccat cttgggtttcc gggacacactt cttcccaagat 660
cggccgggggg ctttggggctt cggcagggggtt gaaaggggcc ctgggtctgg gttcccaaccc 720
55 agtgggtcatg agcacgttgg aggggttaagc cttcaaaatgtt gtcggcaggcc ggggtgtcaga 780
ggtgaagaagat tttccgttca gctttggatctt gggggagggc acatgtgttggaa acccacaagg 840
accttttttctt ctgtacttccat gagct 865
60
<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens
65
<400> 10
tgtggatgtt gttttcatgtt gttggatagg tggggatctgtt gggatgttgggtt tttatgtgtt 60
ggggtaacac agatgttcaag gtcgttccat tttctgtgtt ggggtctgttcat gtcgttccat 120

```

agctttattt aggagaccat atcttccctt gaactatggc cgggttata gtaagtca 180
 ggtgtggagg cctccccctgg gctccctgtt ctgtttcttc cactctgggg tcgtgtgggt 240
 cctgtgtgg tgggtggccg gtgggcaggc cttccaggcc tctttgtgtt cattggccctg 300
 5 gatgtggccc tggctacgtt cctgtgtgg aattccctg cgagtggag gcttttttc 360
 tttttttttt tctttttttt ttttttttt tgataacaga gtcgtctt tttttggcca 420
 ggctggagtg gtttggcggtg atcttggctc actgcacacct gtgttccctg agttcaagca 480
 atctcttgc ctcagccctc caagtagctg gaattatagg cggccaccac catgtgact 540
 aatttttgc attttagtag agacgagggtt tctccatgtt gggccaggctg gtctcaact 600
 10 cctgacccca ggtgtatcc cccacccggc ctcccaaaagt gctggatga cagggtgtgaa 660
 ccggccgcgc cggccggagac tcgcttctgtt cagttccgt gagatctgca gcgatagctg 720
 cctgcagccct tgggtgtgtt aaccctccgtt ttcttctctt aggttccgtt aggggtctt 780
 ccattttcatg actcttccca cagaagagttt tcaatgttgc tgattttccg gcttttctt 840
 gctgttggg tggatgtgtt tttatcgatgg cctccatccca ttttttttag gctttgttta 900
 ttgttggttt tccggcttctt tgaaggaaaa gtttgcattt tggatgtttt aactttttttt 960
 15 tctaaacaag catctgaagt tgcgcgtttt cctctaaaggc agggatcccg aggcccttgg 1020
 ctgtggagtg gcacccgtt gggccctgtt aggaaccccg cgcacagccg gaggcttaggt 1080
 ggggtgtggg gagccagctg tcccccgtt gccccccccc ttcgtatca gcagtggcat 1140
 gctgtgtca gaggcccaaa cccactactt agaactgtgc tgtagggggg tctagatttt 1200
 gtgtcttca tgggaatctt atgcctgtatg atctgggtt gaaagggtt ctccaaaaac 1260
 20 catcccccttc cccactgtt cccatgttcaaaatgttccctt ccacggaaacc agtcccttgg 1320
 accacaatgg ttggggaccc ttttctgtgtt tgatgttcaaaatgttcccttgcgtt 1380
 atatattggc ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 1440
 cggatgttca cccatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1500
 cggggccat gtagccatcc ttttctgtgtt tccatgttcaaaatgttcaaaatgttcaaaatgttca 1560
 25 tccatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1620
 agcacagatg cccatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1680
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1740
 tccctgtgtt aatgacaaggc ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 1800
 tccctgtgtt aatgacaaggc ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 1860
 30 gcaatccctc cccatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1920
 gactgtggat ggcacgttccggtt cccatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 1980
 tcacagggtt ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2040
 gtggatggcg gtcgtggggg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 2100
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2160
 35 ggggtctgtt ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2220
 ggcggccgtt ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2280
 gactgtggat ggcggccgtt ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 2340
 ctgtatgtgtt tgactgtgtt atggccgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2400
 cgggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2460
 40 ctgtgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2520
 ggtctgtatg tggtgtactt ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 2580
 gtggatggcg gtcgtggggg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 2640
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2700
 45 ggtggccgtt ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2760
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2820
 ggggtctgtatg tggtgtactt ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 2880
 ggcggccgtt cccgggggtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 2940
 gtggatggcg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3000
 ggggtctgtatg tggtgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3060
 50 gatggccgtt tggtgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3120
 ggtgtactgtt tggtgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3180
 acagggttctt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3240
 gtggatggcg gtcgtggggg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 3300
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3360
 55 gtcacgggg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3420
 actgtggatg ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3480
 tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3540
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3600
 acgttgcgtt cccggccatgg ggggtctgtt tgatgttcaaaatgttcaaaatgttcaaaatgttca 3660
 ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3720
 60 agtggccatgg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3780
 agtggccatgg ttttctgtgtt tgatgttcaaaatgttcaaaatgttcaaaatgttcaaaatgttca 3782

5 gggattgtcc aatgtggtcc ccctaaggcg cggccacag agccgggg 660
 agtgcgatt gacggggac gagaaaccc taaagctgt aaggaaacc tcagaaaatg 720
 tggccgcac gggtggttc aggtgcattt ctggctgt tttgtaaaaa cccatggaa 780
 cccgcctcc aagtccaccc tccaggtca ccctccaggcg cgcctggg ctgggggtat 840
 gcctggcgct ccttgcgcg cagcccgag cacagcagc tgcacatt taaatccact 900
 aagattcaact cggggggagc ccaggccca agcaactgag ggctcaggag tcctgaggct 960
 gctgaggggcagac gggaaacgct gttctgttgc ggcatttcc tgagggtgct 1020
 ggcaggaggatggctcaga gtgtatgttgc gggtccacc gggggcagaa ctctgttct 1080
 gatgagtggc cagccatgtt acaaggagggttggccacag ggaggctggg atgcaccagg 1140
 10 ggagctgcgcg agctggccga ggtcccgagg ccaggccaca ggaaggggcag ggggacgccc 1200
 gggccacacccagaggccgc aggaaggaa ggggatggccaggccagc agaggctacc 1260
 gggcaggaggatggctccctg agctgggttgc gcgaggctca tgactcggcg agggaccc 1320
 ctgtacgttgc aactgtacatc tgggtgtcc cagtcacag cccagccagg tcccgccct 1380
 gagcaggaaatcagaaccccttgc taaagcacag cagatgcctt cagggcatct 1440
 15 aggagaaaac aggcaaaatc gttgagaaac gtctttaaaag aagggtggat ggtggcaatt 1500
 tcttgcctccat ttttagtcttgc ccccgccacc acagatgttgc tataaaccggg attgtgggttgc 1560
 tgcacatggggg acacatgttgc tggaccatca cagaggccac tgggtgtca cctccatct 1620
 gagtcctggc tggcccgaggatggccatgc ttcacatc tgcctgtcc 1680
 ggggacaggaaatcaggcccaatcaccatctc tgggttttc caaagcattt 1740
 20 cctggccaggcc cagccaccccttgc tcccaatcaccatc tgggttttc caaagcattt 1800
 aacaagggttgc tcaagggttacc tccctgggttgc cggccccc tccctgggttgc gacattgccc 1860
 ctctgccttgc g 1871

25 <210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

30 <400> 15
 gtgagcgcac ctggccggaa gtggagccgt tgcccccgtt gggcagggtgc tgctgcagg 60
 ccgttgcgtc caccctctgttccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggcacagggttgc cccatctggg gctgagcaga aatgcattt tctgtgggag 180
 tgagggtgttgc cacaacggggc gcaatgttgc tgcattttt ggttaaaagga aatgggtcacc 240
 35 cagacccctgg tgcactgttgc aaggatgttgc gatcgaacc caagacgccc 300
 gggccctgttgc gggcggttgc ctctcaaaacc cgaacacagg ggcctgttgc ggcattgttgc 360
 ccttcgttgc cggacccctgg gggccctgttgc gggcggttgc ctctccaaaccc ccaagagactt 420
 caggggccctt tgggtgttgc gtccttcgc tgcgttgc acacttcaag gtcattccac 480
 agtttacaggatgc ttcattgttgc cgtgttgc accaggggac agggccatgg 540
 40 tgggggggg tgccttacaa aatttctggg tctttttccatcccgaggcccg agagctcaag 600
 gcccgttgc accgttgcacaaatgttgc aatgttgc tgcacatc tgcattgttgc 660
 ctgttttctt tataataaaa aatgtatcaatcatttcaggca gggcaagggttgc gtcacaccc 720
 ataatcccaac cactttgggaa gggcgagggttgc ggtggatcac ttggggccatcccg gagtttgggg 780
 ccaacccatcaac caacatagtg aaattccatt tctacttaaa aatataaaaaa attagccctgg 840
 45 ccttgcgttgc caccgttgc tgccttgc tgcggggggc tgaggcgaggaaatcaccatcccg 900
 aaccaggaggatggccatcccg tgcgttgc accgttgcaccc accgttggca 960
 acagagttagtgc accgttgcattt aaaaaaaaaaaa aaaaaaaaaatgttgc caccatcccaaaatcccg 1020
 ggacagggttgc ttttttatttc tgccttgc tataatttttgc tgggtgttgc ttagaggcccg 1080
 gaactgggggg tgccttgc tggaaaggccac accttcatgg gaagagaaaat aatgtgggaa 1140
 tgggtttttaaccatggatgttgc ttcgttgc tgcgttgc accgttgc 1200
 50 tggacttgc ctcttcgttgc aatgttgc tgggttttgc tcatgggggatggccagggttgc 1260
 tggacacccttgc tgggttttgc tggaccccttgc tgcacatc tgcattgttgc 1320
 ggttgcacatc tgggttttgc tgggttttgc tgcacatc tgcattgttgc 1380
 ggatggccgttgc cccatcccg tccatcccg tgcgttgc tgcacatc tgcattgttgc 1440
 ctggcccttcaatcccg tgggttttgc tgcacatc tgcattgttgc 1500
 55 cgcgttgc accgttgc tgggttttgc tgcacatc tgcattgttgc 1560
 ctatccatcccg tgggttttgc tgcacatc tgcattgttgc 1620
 gcaatccatcccg tgggttttgc tgcacatc tgcattgttgc 1680
 gtttgc tgcacatc tgcattgttgc 1740
 60 catcagatgttgc ggggttgc tgggttttgc tgcacatc tgcattgttgc 1800
 tgggtgttgc agagggttgc tgggttttgc tgcacatc tgcattgttgc 1860
 ggggttgc tgggttttgc tgcacatc tgcattgttgc 1920
 ggggttgc tgggttttgc tgcacatc tgcattgttgc 1980
 ttcttgc tgggttttgc tgcacatc tgcattgttgc 2040
 gatggacccatcccg tgggttttgc tgcacatc tgcattgttgc 2100
 65 gatggacccatcccg tgggttttgc tgcacatc tgcattgttgc 2160

5 taggttagact gtcaaatctc agagcaaaat gaaaataaca aagtttaaa gggaaagaaa 2340
aaccggcgtt ctttggactt ctttaggcct gaacttcatt tcaagcagct tccttccaca 2400
gacaaggctg tatggagcga gtgagttcaa agcagaaaagg gaggagaagc aggcagggt 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtgggtttt tcctgcctca 2520
gccccacgtt cctggggcgc ctgcacccgc tgtaaccgtc gatgtgggt ccaggtgccc 2580
acctggaaag gatgctgtgc agggggcttg ccaaactttg gtgggttca gaagccccag 2640
gcacttgcgg caggcacaaat tacagccctt ccccaaaatg gcccacgttcc ttctctggc 2700
acctgtgaat gtgtcaccgg caaggcagag gctggtaag gctgcagggtg gaatcacggc 2760
10 tgccagtcag cccatcttaa ggtcatcttg gattatctgg tggccctgat atggccacaa 2820
gggtccctag aagtggagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
caactggccac tgctggctt gagatggagg agggggccccc cagccaaggaa atgggggcag 2940
ccgcctccatg ctggaaaagc aagcaatctt ccccggtctt gagggcacac ggcctgccc 3000
acgcctcgat ttccaggccag tgggacctgt ttccagtttc cggcctccag agctgttaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattctg cacagcagca aatggaaatag 3120
15 cagtagggg aatgaatac agggacagtt ctcagatgtca ctctcagcccc acccctggg 3179